大学基础数学自学丛书

概率论与数理统计基础

严士健　王隽骧　徐承彝



大学基础数学自学丛书

概率论与数理统计基础

严士健　王隽骧　徐承彝

上海科学技术出版社

大学基础数学自学丛书

概率论与数理统计基础

严士健　王隽骧　徐承彝

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　上海中华印刷厂印刷

开本 787×1092 1/32　印张 12　字数 205,000

1982 年 12 月第 1 版　1982 年 12 月第 1 次印刷

印数 1—52,800

统一书号：13119·1031　　定价：（科三）0.95 元

序　言

我们伟大的祖国，为了尽早实现四个现代化的宏伟大业，需要造就大批又红又专的、具有高度文化修养和现代科学知识的工业大军、农业大军、科技大军、文化大军和国防大军。这是一项摆在全体人民面前的极为艰巨的任务。人才的培养，基础在教育。然而，目前我国每年只可能吸收很小一部分中学毕业生进入高等院校深造，大批已经走上或将要走上各种工作岗位的千千万万青年人，都迫切要求学习现代科学基础知识，以适应新时期的需要。所以，在办好高等院校的同时，还应尽量为那些不能升入大学或无法离职进入大学的青年提供良好的业余学习条件。为此，上海科学技术出版社编辑出版《大学基础数学自学丛书》、《大学基础物理自学丛书》和《大学基础化学自学丛书》。

《大学基础数学自学丛书》由我们负责主编，由北京大学、北京师范大学和复旦大学数学系有关教师执笔编写。包括《一元函数微分学》、《一元函数积分学》、《多元函数微积分》、《级数》、《空间解析几何》、《高等代数》、《复变函数论基础》、《常微分方程基础》、《概率论与数理统计基础》、《微分几何基础》、《有限数学引论》等共十一种，可供具有相当于高中文化程度、有志于自学大学数学课程的广大读者使用。

本《丛书》是一套大学基础课的自学读物，与中学程度的《数理化自学丛书》相衔接。为了使自学读者在没有教师讲课的条件下读懂、学好，其内容选取和编排不同于一般的大学课

本。文字叙述用讲课的形式书写；概念引入尽量从具体的、通俗的地方入手，逐步深入；内容安排抓住重点，讲深讲透。为了对读者解题有所启发，巩固所学的基础知识等，文中举有较多的例题；凡估计读者容易发生困难的地方，尽量给予必要的分析。习题、例题均按章分节安排，书后附有习题答案或提示。每册之首都有编者的话，指导读者自学全书。总之，想尽可能减少自学中的困难。

自学，时间总比在校学习紧得多。要自学有成就，没什么"诀窍"，如果有的话，那就是"多思考，多练习，熟能生巧"。

学习必须从自己的实际水平出发，学每本书要有一定的基础。选读顺序可根据编者的话的指导进行。有志者，事竟成。希望广大读者循序渐进、持之以恒、锲而不舍地学习。愿大家努力学好。

《丛书》编审过程中得到了北京大学数学系、北京师范大学数学系和北京师范学院数学系领导的大力支持；许多同志参加了提纲、样稿的讨论，并提供了宝贵的意见；编撰者和审稿人为《丛书》付出了辛勤的劳动，谨此一并致谢。

由于《丛书》编写和出版的时间仓促，难免有缺点和错误，希望读者不吝赐教！

江泽涵　　　　赵慈庚

于北京大学燕南园　于北京师范大学工五楼

1980年1月

编者的话

本书是概率论和数理统计的一本入门书，内容包括概率论和数理统计的基本概念、结论和方法，为读者进一步学习这方面的知识和进行实际应用提供一个必要的基础．读者只要具有数学分析和线性代数知识就可以阅读本书．

概率论的研究对象是随机现象的数量规律，这和描述决定性现象的数学相比自然有它的特点．为了使初次接触这一数学方法的读者易于领会它的基本概念，本书注意了说明各概念的现实背景和实际含义．我们在不使用更多的数学工具的条件下，努力使叙述和论证保持一定的严谨性和一般性，对不加证明的结论指出了必要的参考读物．为了便于自学，在叙述上我们希望做到简明易懂．但由于作者水平所限，本书离这些目标可能还有不小差距，书中也会有许多不妥和错误，恳请读者提出批评和意见，以便重印时修改．

本书在编写过程中，得到有关兄弟院校的热情关怀和支持，得到教研室其它同志的支持和帮助，杨文礼同志阅读了全书，提出了许多宝贵的意见，谨此一并表示衷心谢意．

编　者
于北京师范大学数学系
1981 年 12 月

目　录

序言

编者的话

第一章　基本概念

第二章　随机变量

第三章 数字特征与特征函数

第四章 大数定律及中心极限定理

第五章 统计推断初步

第一章

基本概念

引言

概率论是数学的一个分支，它研究随机现象中有关事件的规律性.

在一定的条件下必然发生某一事件的现象叫决定性现象.例如纯水在一个大气压力下加热到100°C必然沸腾;质量为m的物体受到外力f的作用必然产生加速度a，等等.

在一定的条件下可能出现这样的结果，也可能出现那样的结果,而且不能预先断言出现哪种结果的现象叫随机现象.例如往桌面上投一枚硬币，可能正面向上，也可能反面向上，而且在投掷前不能预先断言一定哪一面向上；从含有一定个数次品的一批产品中任意取出3件，取到次品的件数可能是0，1，2或3；一门炮对一目标射击,尽管经过瞄准，炮弹却可能落在目标附近的各个位置.

随机现象中的事件在条件实现时可能发生也可能不发生，称在一定条件下可能发生也可能不发生的事件为这个条件下的随机事件. 例如"正面向上"是"往桌面上投一枚硬币"这个条件下的随机事件;"废品数不超过2"是"从含有一定件数次品的一批产品中任意抽取3件"这个条件下的随机事件.称在一定条件下必然发生的事件为这个条件下的必然事件.称在一定条件下必然不发生的事件为这个条件下的不可能事件."水沸腾"是条件——"纯水，一个大气压，100°C" 下的

必然事件，而"水结冰"是上述条件下的不可能事件. 为了方便，把必然事件，不可能事件看作特殊的随机事件.

观察一定条件下发生的现象通常叫做试验；条件实现一次就是一次试验. 一个试验如果可以在相同的条件下重复进行，而且每次试验的结果可以不同，有偶然性，那么我们称它为随机试验. 很明显，随机试验观察的对象实际上就是一个随机现象. 今后，我们把一定条件下的随机现象也称为相应随机试验中的随机现象，而把一定条件下的随机事件称为相应随机试验中的随机事件. 例如，"正面向上"是"投一枚硬币"条件下的随机事件，也可以说成是"投一枚硬币"这一随机试验中的事件. 又如"废品数不超过 2"可以说成是"从含有一定件数废品的一批产品中任意抽取 3 件"这一随机试验中的随机事件.

为了简单起见，今后我们也常常将随机事件简称为事件，将随机试验简称为试验.

随机事件在一次试验中（即条件一次实现之下），可能发生也可能不发生，带有不确定性，但在多次重复试验中，却呈现出明显的规律性.

历史上进行过"投一次硬币"的试验来观察"正面向上"这一事件发生的规律，表 1-1 是试验结果的记录.

表 1-1

实验者	投掷次数	正面向上次数	频率
蒲丰(Buffon)	4040	2048	0.5069
皮尔逊(K. Pearson)	12000	6019	0.5016
皮尔逊(K. Pearson)	24000	12012	0.5005

又如，观察"由 0, 1, 2, …, 9 中任意取出一个数字"的随机试验中，事件"取出的是数字 1"发生的规律. 表 1-2 列出了

10 个 2000 次的观察结果.

表 1-2

观察次数	2000	2000	2000	2000	2000
"1"出现次数	194	203	218	185	212
频率	0.0970	0.1015	0.1090	0.0925	0.1060
观察次数	2000	2000	2000	2000	2000
"1"出现次数	205	183	204	204	205
频率	0.1025	0.0915	0.1020	0.1020	0.1025

注 上表是利用电子计算机进行模拟试验所得的结果.

在观察某一随机事件 A 时, 设进行了 n 次试验, 其中事件 A 发生了 m_A 次, 则 $\frac{m_A}{n}$ 称为事件 A 发生的**频率**.

从上面的试验记录中可以看到, 在多次重复试验中, 同一事件发生的频率虽然并不完全相同, 但却在一个固定的数值附近摆动, 而呈现出一定的稳定性(前一例摆动于 $\frac{1}{2}$ 附近, 后一例摆动于 0.10 附近), 而且随着试验重复次数的增加, 这种现象愈加显著. 频率的这种稳定性, 揭示出一个随机事件发生的可能性有一定的大小可言: 频率稳定于较大数值, 表明相应事件发生的可能性较大, 频率稳定于较小数值, 表明相应的事件发生的可能性较小, 而频率所接近的这个固定数值就是相应事件发生可能性大小的一个客观的定量的度量, 称为相应事件的**概率**.

随机现象广泛地存在于自然界和人类社会中. 随着生产的发展, 科学研究的深入, 许多学科和部门不断提出了研究随机现象的课题, 概率论的方法已广泛地渗透到物理、化学、生

物、地学等学科的研究中和工业、农业、国防等各个部门的工作中．今天，概率论已成为有广泛应用的，有深刻理论基础的，蓬勃发展的一个数学学科了．

第一节　样本空间与事件域

1.1　样本空间

在引言中已经说明，概率论的研究对象是一定条件下的随机现象或随机试验．今后对随机试验理解为：

1．它可以在相同的条件下重复进行；

2．每次试验中可以出现不同的结果，而究竟出现哪一个结果，试验前不能预先断言；

3．试验中一切可能的结果是事先已知的，并且假定在一次试验中必有其中一个结果出现（因为它们是一切可能的结果）且仅有其中一个结果出现（这表示它们之中的任何两个都不能同时发生）．

随机试验中每一个可能的结果称为一个样本点（或基本事件）．全体样本点的集合*)称为样本空间．

在概率论中讨论一个随机试验时，首先要求明确它的样本空间．对于一个具体的随机试验来说，样本空间可以根据试验的内容（即试验条件实现一次的含义和观察的目的）来决定．

【例1】　投一枚硬币观察正反面出现情况．一次试验就是投一枚硬币，试验的结果有两个：正（正面向上），反（反面向

*) 在以后的叙述中，将用到集合的概念和运算，不熟悉这些基本知识的读者可参考本丛书《有限数学引论》．

上)，即有两个样本点：正、反，这个随机试验的样本空间为

$$\Omega=\{正、反\}.$$

【例 2】 将一枚硬币投两次，观察正反面出现情况．在这个随机试验中，投两次硬币是一次试验（注意：不是两次试验！），试验结果有 4 个：(正，正)，(正，反)，(反，正)，(反，反)．这里的记号，如(正，反)，表示"第一次出现正面，第二次出现反面"，其余类此．样本空间

$$\Omega=\{(正,正),(正,反),(反,正),(反,反)\}.$$

【例 3】 从包含两件次品（记作 a_1, a_2）和三件正品（记作 b_1, b_2, b_3）的五件产品中，任意取出两件．具体拿出两件就是一次试验，例如拿出的两件是 a_1 和 b_3，这就是一个样本点，记作 $\{a_1, b_3\}$．所有样本点共 10 个，样本空间

$$\Omega=\{\{a_1, a_2\}, \{a_1, b_1\}, \{a_1, b_2\}, \{a_1, b_3\}, \{a_2, b_1\}, \{a_2, b_2\}, \{a_2, b_3\}, \{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}.$$

【例 4】 把红、绿、白三个球随机地放到 I、II、III 三个盒中（每盒可以容纳任意多个球），这个随机试验的样本点共有 27 个，分别记作 ω_1, ω_2, $\cdots$, ω_{27}，图示如下（a, b, c 分别表示红、绿、白三个球，I、II、III 三个盒从左向右排列）：

ω_1	abc			ω_{10}	bc	a		ω_{19}	bc		a
ω_2	ab	c		ω_{11}	b	ac		ω_{20}	b	c	a
ω_3	ab		c	ω_{12}	b	a	c	ω_{21}	b		ac
ω_4	ac	b		ω_{13}	c	ab		ω_{22}	c	b	a
ω_5	a	bc		ω_{14}		abc		ω_{23}		bc	a
ω_6	a	b	c	ω_{15}		ab	c	ω_{24}		b	ac
ω_7	ac		b	ω_{16}	c	a	b	ω_{25}	c		ab
ω_8	a	c	b	ω_{17}		ac	b	ω_{26}		c	ab
ω_9	a		bc	ω_{18}		a	bc	ω_{27}			abc

【例 5】 记录某电话交换台一分钟内接到的呼叫次数. 样本点(记录结果)是一非负整数(接到的呼叫数), 由于难于规定一个呼叫数的上界, 所以认为每一个非负整数都是一个可能结果, 故样本空间 $\Omega=\{0, 1, 2, \cdots\}$.

【例 6】 向某一目标射击一发炮弹, 观察落点与目标的偏差. 样本点可以是任何一个非负实数(偏差值), 所以样本空间 $\Omega=\{d \mid d\geqslant 0\}$.

【例 7】 向某一目标射击一发炮弹, 观察落点的分布情况. 每次试验结果就是"炮弹落在某一平面区域 G 的某一点 (x, y) 上(假设平面已经建立坐标系). 这个结果可以简单地就用点 (x, y) 来表示. 于是样本点就是 G 的各个点, 样本空间 $\Omega=G$.

从上面的例中我们看到为了给出样本空间, 必须确切理解试验条件的内容和观察的目的, 比较例 1 和例 2, 例 6 和例 7, 可以使我们更看清这个问题.

1.2 事件

随机试验中的事件, 通常用语言来表述, 引入了样本空间之后, 事件便可以定义为样本空间的子集. 先结合本节例 3 来理解. 考虑事件 A_0: "没有抽到次品". 在一次试验中, A_0 发生, 当且仅当在这次试验中出现样本点 $\{b_1, b_2\}$, $\{b_1, b_3\}$, $\{b_2, b_3\}$ 中的一个. 这样我们可以认为 A_0 是由 $\{b_1, b_2\}$, $\{b_1, b_3\}$, $\{b_2, b_3\}$ 组成的, 而将 A_0 定义为它们组成的集合,

$$A_0=\{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}.$$

又如事件 A_1 "抽到一个次品". A_1 发生当且仅当 $\{a_1, b_1\}$, $\{a_1, b_2\}$, $\{a_1, b_3\}$, $\{a_2, b_1\}$, $\{a_2, b_2\}$, $\{a_2, b_3\}$ 中的一个发生, 于是我们可以认为 A_1 是它们组成的而定义为它们组成的集

合：

$$A_1=\{\{a_1,\ b_1\},\ \{a_1,\ b_2\},\ \{a_1,\ b_3\},\ \{a_2,\ b_1\},\ \{a_2,\ b_2\},\ \{a_2,\ b_3\}\}.$$

又如事件 A_2："抽到两件次品"．A_2 发生当且仅当样本点 $\{a_1,\ a_2\}$ 发生．将 A_2 定义为 $\{a_1,\ a_2\}$ 组成的单元素集：$A_2=\{\{a_1,\ a_2\}\}$．

一般，我们将事件定义为样本空间 Ω 的某个子集．称事件 A 发生当且仅当 A 中某一样本点发生．

再看几个例．

在例 2 中：事件 A"两次出现的面不同"，

$$A=\{(\text{反},\text{正}),(\text{正},\text{反})\}.$$

事件 B"第一次出现反面"，

$$B=\{(\text{反},\text{正}),(\text{反},\text{反})\}.$$

在例 6 中：事件 A"偏差不超过 100 米"，

$$A=\{d\,|\,0\leqslant d\leqslant 100\}.$$

在例 7 中：事件 A"炮弹落在目标(0, 0)的左侧"，

$$A=\{(x,\ y)\,|\,(x,\ y)\in G,\ x<0\}.$$

事件 B"炮弹落在距目标不超过 100 米的范围"，

$$B=\{(x,\ y)\,|\,x^2+y^2\leqslant 100^2\}.$$

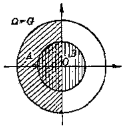


图 1-1

样本空间 Ω 作为一个事件，因为在每次试验中必有 Ω 中的样本点出现，所以 Ω 在一次试验中必然发生，即 Ω 是必然事件．

空集 $\emptyset$ 作为一个事件，它在每一次试验中都不会发生，所以 $\emptyset$ 是不可能事件．

1.3 事件的关系与运算

现在，对应着集合的关系和运算，定义事件的关系及运算。

在以下的叙述中，设 Ω 是一给定的样本空间，所出现的 Ω 的子集都是事件，并用 $A, B, C, \cdots$ 表示。

1．若 $A\subseteq B$，则称事件 B 包含事件 A。

事件 B 包含事件 A 的含义是：在一次试验中，若 A 发生则 B 必发生．这是因为，A 发生，一定有一个样本点 ω 出现，$\omega\in A$，由于 $A\subseteq B$，因此 $\omega\in B$，所以 B 发生。

例如，在例 4 中，

$$A=\{\omega_1, \omega_2, \cdots, \omega_9\}，“红球在 I 盒中”；$$

$$B=\{\omega_1, \omega_2, \cdots, \omega_{18}\}，“红球不在 III 盒中”。$$

显然 $A\subseteq B$，而从实际含义来看就是：若事件“红球在 I 盒中”发生，则事件“红球不在 III 盒中”必发生。

2．若 $A=B$，则称事件 A 与事件 B **相等**。

$A=B$ 表示它们是同一个事件。

3．事件 $A\cup B$ 称为事件 A 与事件 B **之和**。

在一次试验中，$A\cup B$ 发生，当且仅当出现某一样本点 $\omega\in A\cup B$(即 $\omega\in A$ 或 $\omega\in B$)，当且仅当 A 发生或 B 发生[*]。由此可见，用语言表述时 $A\cup B$ 就是“A 或 B”这一事件，通常也把这个事件表述为“A 发生或 B 发生”，也表述为“A, B 至少有一个发生”。

例如，在例 4 中

$$A=\{\omega_1, \omega_2, \cdots, \omega_9\}，“红球在 I 盒中”；$$

---

[*] “A 发生或 B 发生”的含意包括 A, B 同时发生。

$B=\{\omega_{19}, \omega_{20}, \cdots, \omega_{27}\}$，"红球在 III 盒中"。

$A\cup B=\{\omega_1, \cdots, \omega_9, \omega_{19}, \cdots, \omega_{27}\}$，

"红球在 I 盒中或红球在 III 盒中"。

4．事件 $A\cap B$ 称为事件 A 与事件 B 的交（A，B 的交也记作 AB）。

$A\cap B$ 发生当且仅当 A 发生且 B 发生．$A\cap B$ 就是"A 且 B"这一事件，也常常表述为"A 发生且 B 发生"或"A，B 同时发生"等等．

例如，在例 4 中

$A=\{\omega_1, \cdots, \omega_{18}\}$，"红球不在 III 盒中"；

$B=\{\omega_{10}, \cdots, \omega_{27}\}$，"红球不在 I 盒中"。

$A\cap B=\{\omega_{10},\cdots,\omega_{18}\}$，"红球不在 III 盒且不在 I 盒中"

（即"红球在 II 盒中"）．

5．事件 $A-B$ 称为事件 A 与事件 B 之差．

事件 $A-B$ 发生，当且仅当 A 发生且 B 不发生．用语言表述时，$A-B$ 就是"A 但非 B"这一事件．通常也表述为："A 发生且 B 不发生"．

例如，在例 4 中

$A=\{\omega_1, \cdots, \omega_9\}$，"红球在 I 盒中"；

$B=\{\omega_1, \omega_2, \omega_3, \omega_{10}, \omega_{11}, \omega_{12}, \omega_{19}, \omega_{20}, \omega_{21}\}$，

"绿球在 I 盒中"。

$A-B=\{\omega_4, \omega_5, \cdots, \omega_9\}$，"红球在 I 盒中且绿球不在 I 盒中"．

6．事件 $\overline{A}=\Omega-A$ 称为事件 A 的对立事件．

$\overline{A}$ 发生当且仅当 A 不发生．用语言表述时 $\overline{A}$ 就是"非

A" 这一事件，通常也表述为"A 不发生".

例如，在例 5 中

$A=\{k|k>100\}$，"接到的呼叫超过 100 次".

$\overline{A}=\{k|k\leqslant 100\}$，"接到的呼叫不超过 100 次".

由集合的运算性质，$\overline{(\overline{A})}=A$，即 A 是 $\overline{A}$ 的对立事件，A 也是 $\overline{A}$ 的对立事件.

由集合的性质还知，$A\cup\overline{A}=\Omega$，$A\cap\overline{A}=\emptyset$. 这说明，在任何一次试验中相互对立的事件 A 与 $\overline{A}$，必有一个且仅有一个发生.

7．如果 $A\cap B=\emptyset$，则称事件 A，B 互不相容.

互不相容的事件就是不能在同一次试验中发生的事件. 通常也把互不相容的事件表述为不能同时发生的事件.

例如，在例 5 中

$A=\{k|k>100\}$，"接到的呼叫超过 100 次"；

$B=\{k|50\leqslant k\leqslant 100\}$，"接到的呼叫在 50 到 100 次之间".

$A\cap B=\emptyset$，故 A，B 互不相容.

相互对立的事件一定是互不相容的，但反过来一般并不成立. 上面的 A，B 互不相容，但 $B\neq\overline{A}$.

事件和、交的概念可以推广到有限个或无限多个事件. 设 $A_i\subseteq\Omega(i=1,\ 2,\ \cdots)$是给定的事件.

称 $\bigcup\limits_{i=1}^{n}A_i$ 为事件 A_1，$\cdots$，A_n 的和，通常表述为"A_1，$\cdots$，A_n 至少有一个发生".

称 $\bigcup\limits_{i=1}^{\infty}A_i$ 为事件 A_1，$\cdots$，A_n，$\cdots$ 的和，通常表述为"A_1，$\cdots$，A_n，$\cdots$ 至少有一个发生".

事件的和也常常用符号 $A_1+\cdots+A_n$；$A_1+\cdots+A_n+\cdots$

以及 $\sum_{i=1}^{n} A_i$, $\sum_{i=1}^{\infty} A_i$ 来表示.

称 $\bigcap_{i=1}^{n} A_n$ 为事件 $A_1, \cdots, A_n$ 的交, 通常表述为: "$A_1, \cdots, A_n$ 同时发生".

称 $\bigcap_{i=1}^{\infty} A_n$ 为事件 $A_1, \cdots, A_n, \cdots$ 的交, 通常表述为: "$A_1, \cdots, A_n, \cdots$ 同时发生".

事件的交也常常称为事件的积, 并用符号 $A_1 \cdots A_n$, $A_1 \cdots A_n \cdots$, 以及 $\prod_{i=1}^{n} A_i$, $\prod_{i=1}^{\infty} A_i$ 来表示.

上面我们利用集合论的概念引入了概率论的一些基本概念, 为了便于对照列成下表:

表 1-3

符号	概率论	集合论
Ω	样本空间 必然事件	空间(全集)
$\emptyset$	不可能事件	空集
ω	样本点(基本事件)	元素
A	事件	子集
$\bar{A}$	A 的对立事件	A 的余集
$A \subseteq B$	事件 A 发生必有事件 B 发生	A 是 B 的子集
$A = B$	事件 A 与事件 B 相等	A 与 B 相等
$A \cup B$	事件"A 与 B 至少有一个发生"	A 与 B 的和集
AB	事件"A 与 B 同时发生"	A 与 B 的交集
$A - B$	事件"A 发生而 B 不发生"	A 与 B 的差集
$AB = \emptyset$	事件 A 和事件 B 互不相容	A 与 B 没有公共元素

事件的关系和运算也常用图形来直观示意:

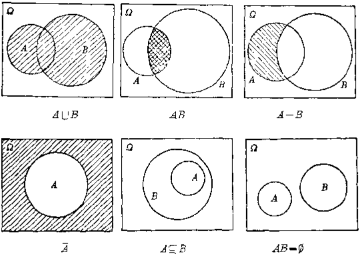


图 1-2

关于事件的运算有下列的基本关系式:

交换律　$A\cup B=B\cup A$, $AB=BA$;

结合律　$(A\cup B)\cup C=A\cup(B\cup C)$, $(AB)C=A(BC)$;

分配律　$(A\cup B)C=AC\cup BC$,

$$AB\cup C=(A\cup C)(B\cup C);$$

德莫根　(DeMorgan)公式(对偶公式)

$$\overline{A\cup B}=\bar{A}\bar{B},\ \overline{AB}=\bar{A}\cup\bar{B}.$$

分配律和德莫根公式可以推广到任意有限个事件或无限个事件,如

$$(\bigcup_i A_i)C=\bigcup_i A_iC;$$

$$\overline{\bigcup_i A_i}=\bigcap_i \bar{A}_i,\ \overline{\bigcap_i A_i}=\bigcup_i \bar{A}_i.$$

我们常常需要利用给定的事件通过运算表出另外一些事件.

【例 8】 设 A_1, A_2, A_3, A_4 是四个事件，试用它们表出下列事件: B:"A_1, A_2, A_3, A_4 恰有一个发生", C:"A_1, A_2, A_3, A_4 至多有一个发生".

解: B="恰好 A_1 发生, 或恰好 A_2 发生, 或恰好 A_3 发生或恰好 A_4 发生"

而"恰好 A_1 发生"="A_1 发生, 且 A_2 不发生,

且 A_3 不发生, 且 A_4 不发生"

$=A_1\overline{A}_2\overline{A}_3\overline{A}_4.$

类似地 "恰好 A_2 发生"$=\overline{A}_1A_2\overline{A}_3\overline{A}_4,$

"恰好 A_3 发生"$=\overline{A}_1\overline{A}_2A_3\overline{A}_4,$

"恰好 A_4 发生"$=\overline{A}_1\overline{A}_2\overline{A}_3A_4.$

所以

$$B=A_1\overline{A}_2\overline{A}_3\overline{A}_4+\overline{A}_1A_2\overline{A}_3\overline{A}_4+\overline{A}_1\overline{A}_2A_3\overline{A}_4+\overline{A}_1\overline{A}_2\overline{A}_3A_4,$$

C="A_1, A_2, A_3, A_4 都不发生或恰有一个发生"

$$=\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4+B$$

$$=A_1\overline{A}_2\overline{A}_3\overline{A}_4+A_1\overline{A}_2\overline{A}_3\overline{A}_4+\overline{A}_1A_2\overline{A}_3\overline{A}_4+\overline{A}_1\overline{A}_2A_3\overline{A}_4+\overline{A}_1\overline{A}_2\overline{A}_3A_4.$$

【例 9】 如图1-3有一电路. A_i 表示事件"继电器接点 i 闭合". B 表示事件: "R, S 两点间是通路".

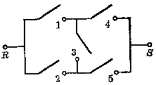


图 1 3

解: R, S 间是通路当且仅当"1, 4 同时闭合"或"1, 3, 5 同时闭合"或"2, 3, 4 同时闭

合"或"2, 5同时闭合". 所以

$$B=A_1A_4+A_1A_3A_5+A_2A_3A_4+A_2A_5.$$

1.4 事件域

在讨论一个随机现象时, 与所要考虑的问题有关的常常不止一个事件而是若干个事件, 为了讨论这些事件又涉及到它们的和、交、差及对立事件, 等等. 另一方面, 一般说来, 又不能把样本空间 Ω 的所有的子集都作为事件包括在我们的讨论范围之内, 那样将会给进一步的讨论带来困难. 通常的作法是, 根据具体情况确定一组事件, 记它们组成的集合为 $\mathscr{F}$, 要求 $\mathscr{F}$ 包含与问题有关的事件, 而又对和、交、差、余运算封闭.

定义 由样本空间 Ω 的一些子集组成的集合 $\mathscr{F}$, 叫做 Ω 上的一个 σ-域或事件域, 如果满足下列条件:

1) $\Omega\in\mathscr{F}$.

2) 若 $A\in\mathscr{F}$, 则 $\overline{A}\in\mathscr{F}$.

3) 若 $A_i\in\mathscr{F}(i=1, 2, \cdots)$, 则 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathscr{F}$.

称 $\mathscr{F}$ 中的元素为事件.

事件域的定义中只要求对求余及求和封闭, 下面我们证明, 事件域对求积, 求差等运算也是封闭的.

1′) $\emptyset\in\mathscr{F}$.

【证】 由 1), $\Omega\in\mathscr{F}$, 再由 2), $\emptyset=\overline{\Omega}\in\mathscr{F}$. ▌

2′) 若 $A_i\in\mathscr{F}(i=1, 2, \cdots)$, 则 $\bigcap\limits_{i=1}^{\infty}A_i\in\mathscr{F}$.

【证】 由 2), $\overline{A}_i\in\mathscr{F}$ $(i=1, 2, \cdots)$, 由 3), $\bigcup\limits_{i=1}^{\infty}\overline{A}_i\in\mathscr{F}$, 利用对偶公式及 2)

$$\bigcap_{i=1}^{\infty} A_i = \overline{\left(\bigcup_{i=1}^{\infty} \overline{A}_i\right)} \in \mathscr{F}. \quad ▌$$

3′) 若 $A_1, \cdots, A_n \in \mathscr{F}$, 则 $A_1 \cup \cdots \cup A_n \in \mathscr{F}$.

4′) 若 $A_1, \cdots, A_n \in \mathscr{F}$, 则 $A_1 \cap \cdots \cap A_n \in \mathscr{F}$.

事实上,在 3) 中取 $A_i = \emptyset (i = n+1, n+2, \cdots)$ 即得 3′). 在 4)中取 $A_i = \Omega (i = n+1, n+2, \cdots)$ 即得 4′).

5′) 若 $A, B \in \mathscr{F}$, 则 $A - B = A\overline{B} \in \mathscr{F}$.

我们举几个事件域的例.

【例 10】 $\mathscr{F} = \{\emptyset, A, \overline{A}, \Omega\}$ 是一个事件域.

【例 11】 $\mathscr{F} = \{\Omega$ 的一切子集$\}$是一个事件域.

当 $\Omega = \{\omega_1, \cdots, \omega_n\}$ 是一有限集时, $\mathscr{F}$ 包含 2^n 个元素. 如果 Ω 包含不可数无限多个元素时,这个 $\mathscr{F}$ 就非常复杂.

上面我们提出取一个适当的事件域 $\mathscr{F}$ 作为我们所讨论的事件的一个范围. 然而究竟怎样的事件域才是适当的,对这个问题要想作出一般回答是困难的. 现在我们针对一些常见的情形作一些说明.

离散样本空间

如果样本空间 Ω 只有有限个元素或可数多个元素 (即它的全体元素可以排成一个序列 $\omega_1, \omega_2, \cdots, \omega_n, \cdots$), 则称 Ω 是离散的. 对于离散的样本空间, 一般总取 Ω 的全体子集组成的集合 $\mathscr{F}$ 作为事件域.

波雷尔(Borel)域

如果 Ω 是一实数区间(有限的, 或无限的), 则通常取 $\mathscr{F}$ 为包含 Ω 的一切子区间的一个最小的 σ 域. 这个 σ 域叫作 Ω 上的波雷尔域,其中的集合叫做波雷尔集.

如果 Ω 是二维平面上的矩形(有限或无限),则通常取 $\mathscr{F}$ 为包含 Ω 的一切子矩形的一个最小 σ 域. 这个 σ 域叫作 Ω 上

的波雷尔域. 其中的集合叫做波雷尔集.

多维空间的情形可类似地处理. 详细地讨论超出了本书的范围,这里只能粗浅地提一提这方面的概念.

当 Ω 为上述情形时, 通常总是取波雷尔域作为事件域.

习 题 一

1. 给出随机试验:"从包含两件次品 a_1, a_2 和三件正品 b_1, b_2, b_3 的五件产品中依次取出两件"的样本空间, 并写出下列事件的集合表示:

   A_0: "没有取到次品",

   A_i: "恰好取到 i 件次品", $(i=1, 2)$

   B_1: "第一次取到的是次品".

2. 给出随机试验:"从包含两件次品 a_1, a_2 和三件正品 b_1, b_2, b_3 的五件产品中依次取出两件(每次取出后放回)"的样本空间, 并写出下列事件的集合表示:

   A_0: "没有取到次品",

   A_i: "恰好取到 i 件次品", $(i=1, 2)$

   B_1: "第一次取到的是次品",

   C: "两次取出的是同一件产品".

3. 一口袋装有五张卡片分别写着数码 1, 2, 3, 4, 5, 从口袋中任意依次取出两张组成一个二位数. 试写出这一随机试验的样本空间及下列事件的集合表示:

   A: "组成的数是一偶数",

   B: "组成的数含有数字 2",

   C: "组成的数含有数字 4".

   下列关系哪一个正确: $A\subseteq B$, B、C 互不相容,

   $$A=B\cup C, \quad B\cap C\subseteq A.$$

   如果每次取出卡片后放回,答案如何?

4. A_i 表示"电话交换台在一分钟内接到 i 次呼唤" ($i=0, 1, 2, 3, 4, 5$), A 表示"在一分钟内接到不多于 5 次呼唤", B 表示"在一分钟内接到多于 5 次呼唤". 以上八个事件哪些事件有包含关系? 哪些事件有对立关系? 哪些事件互不相容?

5. 设 A_1 表示"第一次射击击中靶子", A_2 表示"第二次射击击中靶子", A_3 表示"第三次射击击中靶子".

(1) 试用语言表述下列各事件:

$$\overline{A}_1+\overline{A}_2+\overline{A}_3,\quad \overline{A_1+A_2},\quad A_1A_2\overline{A}_3+\overline{A}_1A_2A_3;$$

(2) 试用 A_1, A_2, A_3 通过运算关系表出下列事件: "三次射击中恰好有一次击中靶子", "三次射击中第一次不中而后两次至少有一次击中".

6. 如图 1-4, 1-5 有两个电路. 令 A_i 表示"第 i 接点开关闭合" ($i=1, 2, 3, 4, 5, 6$). 试用 A_i 表出事件:

B_1: "L_1, R_1 是通路",

B_2: "L_2, R_2 是通路".

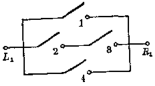


图 1-4

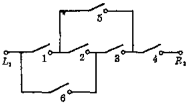


图 1-5

7. 设 A, B, C 是三个随机事件, 试用 A, B, C 表示下列各事件:

(1) 恰有 A 发生;

(2) A 和 B 都发生而 C 不发生;

(3) 所有这三个事件都发生;

(4) A, B, C 至少有一个发生;

(5) 至少有两个事件发生;

(6) 恰有一个事件发生;

(7) 恰有两个事件发生;

(8) 不多于一个事件发生;

(9) 不多于两个事件发生;

(10) 三个事件都不发生.

第二节 概率的定义

在引言中, 我们已经分析了概率概念的现实意义. 一个事件 A 的概率 $P(A)$ 是标志事件 A 发生可能性大小的一个客观存在的数. 在重复试验中, 事件 A 发生的频率在 $P(A)$ 的附近摆动, 因此, 频率可以作为 $P(A)$ 的一个估计值. 然而概率究竟是什么还需要给出明确的定义.

概率的定义曾是概率论发展过程中人们长期探讨的课题. 随着数学工具的发展及对概率概念认识的加深, 而提出了公理化定义, 它的要点可简括如下.

首先, 在一个随机试验中, 我们所要讨论的不仅仅是个别的事件, 而是一个事件域 $\mathscr{F}$ 中的所有事件的概率. 这就是说, 对于每一事件 $A\in\mathscr{F}$, 相对应的 $P(A)$ 都是我们的研究对象. 明确了这一点就使我们认识到, 概率是一个函数, 它的定义域是 $\mathscr{F}$ (即"自变量"是 $\mathscr{F}$ 中的事件).

其次, 通过频率与概率的关系, 我们看到 $P(A)$ 的值都应在 $[0, 1]$ 之间, 所以, 概率是一个在 $[0, 1]$ 中取值的函数(即它的值域为 $[0, 1]$). 并且, 必然事件的概率为 1, $P(\Omega)=1$ (因为必然事件的频率永远为 1).

此外, 通过对频率的考虑, 概率还应该具有可加性. 设 A, B 是两个互不相容的事件. 假设进行了 n 次试验, 在这 n 次试验中, A 发生了 m_A 次, B 发生了 m_B 次. 由于 A, B 互不相容, 可知 $A\cup B$ 发生了 m_A+m_B 次. 于是

$A\cup B$ 发生的频率

$$=\frac{m_A+m_B}{n}=\frac{m_A}{n}+\frac{m_B}{n}$$

$=A$ 发生的频率 $+B$ 发生的频率.

所以，应有

$$P(A\cup B)=P(A)+P(B)\quad (\text{当 } A,\ B \text{ 互不相容}).$$

这就是所谓的可加性. 同理，对于任意有限个或可数个两两互不相容的事件来说可加性也应成立.

正是上面这些认识概括了概率概念的本质. 下面是概率公理化定义的确切叙述.

定义 设 Ω 是一给定的样本空间，$\mathscr{F}$ 是 Ω 中的一个事件域. 定义在 $\mathscr{F}$ 上的一个实值函数 $P(A)(A\in\mathscr{F})$，如果满足下列条件，就称它是 $\mathscr{F}$ 上的一个*概率测度*，简称*概率*，值 $P(A)$ 称为*事件 A 的概率*：

1) 对每个 $A\in\mathscr{F}$，$0\leqslant P(A)\leqslant 1$.

2) $P(\Omega)=1$.

3) 若 $A_i\in\mathscr{F}(i=1,2,\cdots)$ 两两不相容 $(A_iA_j=\emptyset,\ i\neq j)$，则

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i).$$

从定义立刻可以看出概率具有下列性质：

4) $P(\emptyset)=0$.

5) 若 $A_1,\cdots,A_m\in\mathscr{F}$，两两不相容 $(A_iA_j=\emptyset,\ i\neq j)$，则

$$P(A_1\cup\cdots\cup A_m)=P(A_1)+\cdots+P(A_m).$$

4)的证明：在 3) 中取 $A_i=\emptyset(i=1,2,\cdots)$，则有

$$P(\emptyset)=P(\emptyset)+P(\emptyset)+\cdots+P(\emptyset)+\cdots.$$

而由 1)，$P(\emptyset)$ 是非负实数，所以必有 $P(\emptyset)=0$.

5) 的证明：在 3) 中取 $A_i=\emptyset(i=m+1,m+2,\cdots)$，注

意到 $A_1\cup\cdots\cup A_m=A_1\cup\cdots\cup A_m\cup\emptyset\cup\emptyset\cup\cdots$，并利用 4) 即得

$$\begin{aligned}P(A_1\cup\cdots\cup A_m)&\\=P(A_1)+\cdots+P(A_m)+P(\emptyset)&\\+P(\emptyset)+\cdots&\\=P(A_1)+\cdots+P(A_m).&\end{aligned}$$

概率的性质 3) 称为 σ-可加性，性质 5) 称为有限可加性。概率还有许多重要性质将在第五节中继续讨论。

至此，我们已经引入了概率论中最基本的三个概念：样本空间 Ω，事件域 $\mathscr{F}$ 和概率 P。在分析一个随机试验时，我们应该明确它的样本空间，指出所讨论的事件的范围——事件域 $\mathscr{F}$，然后确定出概率 P。由此看来，Ω, $\mathscr{F}$, P 这三者是描述一个随机试验的三个基本组成部分，我们把三者合在一起，称 $(\Omega, \mathscr{F}, P)$ 为一个概率空间。

这样，从概念上讲，研究一个具体问题，应该从确定它的概率空间为起点。也有许多场合，是在假定概率空间存在或已给定的情况下来进行概率计算或其它性质的研究。

在概率的定义中，只指明它是满足条件 1)～3) 的一个 $\mathscr{F}$ 上的函数，并没有指出它的具体形式和事件 A 的概率 $P(A)$ 的具体数值。不同的随机试验有不同的特点，因此，必需联系随机试验的特点才能确定出概率的具体形式和数值。下面两节和第八、九两节，我们将对一定类型的随机试验，确定出它们的概率。

第三节　古 典 概 型

有许多随机试验满足下述两个条件：

（i）它的样本空间只有有限个样本点；

(ii) 每个样本点出现的可能性相同.

这种随机试验是概率论发展早期的研究对象，称为古典型随机试验，简称古典概型. 现在，我们要来确定古典概型中事件的概率 P.

设样本空间为 $\Omega=\{\omega_1, \cdots, \omega_n\}$. 如果每个样本点出现的可能性相同，那么就有

$$P(\{\omega_1\})=\cdots=P(\{\omega_n\}). \tag{1}$$

根据概率的定义，

$$1=P(\Omega)=P(\{\omega_1\})+\cdots+P(\{\omega_n\})=nP(\{\omega_i\}).$$

所以，

$$P(\{\omega_i\})=\frac{1}{n} \quad (i=1, \cdots, n). \tag{2}$$

如果事件 A 包含 m 个样本点:

$$A=\{\omega_{i_1}, \cdots, \omega_{i_m}\}.$$

则由可加性 $\quad P(A)=P(\{\omega_{i_1}\})+\cdots+P(\{\omega_{i_m}\})=\dfrac{m}{n}$,

即

$$P(A)=\frac{A\text{ 中包含的样本点的个数}}{\text{样本点总数}}. \tag{3}$$

以上的讨论说明: 如果在 $\mathscr{F}$ 上存在概率 P 满足 (1), 则事件 A 的概率 $P(A)$ 应该按(3)来确定.

现在我们来进一步证明，由(3)及 $P(\emptyset)=0$ 定义的 $\mathscr{F}$ 上的函数 $P(A)$ 确是一个概率测度并且满足等可能性条件 (1). 为此，需要验证，$P(A)$满足概率的定义. 1), 2)显然满足，主要需验证可加性 3). 对于任意两个互不相容的事件 A, B,

$$\begin{aligned} P(A\cup B) &= \frac{A\cup B\text{ 中样本点个数}}{n} \\ &= \frac{A\text{ 中样本点个数}+B\text{ 中样本点个数}}{n} \\ &= P(A)+P(B). \end{aligned}$$

完全类似地，也不难验证：对于任意有限个两两互不相容的事件来说，可加性成立．由于 $\mathscr{F}$ 中只有有限多个事件，σ-可加性可由有限可加性导出．事实上，若 $A_1, \cdots, A_m, \cdots$ 是两两互不相容的事件，必有某一 m_0，使得 $A_{m_0+1}=A_{m_0+2}=\cdots=\emptyset$，而 $P(\emptyset)=0$，从而有

$$\begin{aligned}P(A_1\cup\cdots\cup A_m\cup\cdots)&=P(A_1\cup\cdots\cup A_{m_0})\\&=P(A_1)+\cdots+P(A_{m_0})\\&=P(A_1)+\cdots+P(A_{m_0})+P(A_{m_0+1})+\cdots.\end{aligned}$$

这样，我们就得出结论，在古典概型中，事件 A 的概率由(3)来确定．

我们来看几个简单的例．

在第一节例 1 中，投一枚硬币的试验有两个样本点：正、反，由于硬币的对称性，两个样本点的出现是等可能的，所以是古典概型．事件 $A=\{$正$\}$（"出现正面"）包含一个样本点，$P(A)=1/2$．

在第一节例 2 中，投两次硬币的试验，样本空间由 4 个等可能的样本点组成，是古典型随机试验．所以有

$A_0=\{$(反，反)$\}$（"正面不出现"），

$$P(A_0)=\frac{1}{4};$$

$A_1=\{$(正，反)，(反，正)$\}$（"出现一次正面"），

$$P(A_1)=\frac{2}{4};$$

$A_2=\{$(正，正)$\}$（"出现两次正面"），

$$P(A_2)=\frac{1}{4};$$

$B=\{$(正，正)，(正，反)$\}$（"第一次出现正面"），

$$P(B)=\frac{2}{4},$$

等等.

在第一节例 3 中,从五件产品中任意取出两件,每一种取法可能性都相同,所以是古典概型. 样本点共 $C_5^2=10$ 个.

$A_0=\{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}$ ("没有取到次品")包含 $C_3^2=3$ 个样本点;

$A_1=\{\{a_1, b_1\}, \{a_1, b_2\}, \{a_1, b_3\}, \{a_2, b_1\}, \{a_2, b_2\}, \{a_2, b_3\}\}$ ("恰好取到一个次品")包含 $C_2^1C_3^1=6$ 个样本点;

$A_2=\{\{a_1, a_2\}\}$ ("恰好取到两个次品") 包含 $C_2^2=1$ 个样本点.

所以, $P(A_0)=\frac{3}{10}$, $P(A_1)=\frac{6}{10}$, $P(A_2)=\frac{1}{10}$. 又事件 B "取到的次品数不超过 1" 是 A_0, A_1 之和, $B=A_0+A_1$, 且 A_0、A_1 互不相容,所以

$$P(B)=P(A_0)+P(A_1)=\frac{3}{10}+\frac{6}{10}=\frac{9}{10}.$$

在第一节例 4 中,将红、绿、白三球任意放到三个盒中,每一种放法都是等可能的,是古典概型. 样本点共 27 个.

$A=\{\omega_1, \cdots, \omega_9\}$ ("红球在 I 盒中"), $P(A)=\frac{9}{27}$.

$B=\{\omega_6, \omega_8, \omega_{12}, \omega_{16}, \omega_{20}, \omega_{22}\}$ ("每盒各有一球"), $P(B)=\frac{6}{27}$, 等等.

从这些例子来看, 计算古典概型中事件 A 的概率, 在原则上并不困难. 可以把样本点全列出来, 数出样本点的总数 n, 再数出事件 A 包含的样本点的个数 m, 就得到 $P(A)=m/n$. 但在样本点较多或问题比较复杂时, 把样本点全列出

来直接去数实际上是不容易的，这时需要利用排列与组合的知识.

【例 1】 180 只集成电路中，如果 172 只是合格品，8 只是次品，试求从中任意抽取 4 只，4 只中"没有次品"（事件 A_0）和恰有"1 只"、"2 只"，"3 只"，"4 只"次品（分别记为事件 A_1, A_2, A_3, A_4）的概率各是多少？"次品超过 1 只"（事件 B）的概率是多少？

解： 从 180 只集成电路中任意抽取 4 只，每种取法是等可能的. 利用组合公式，样本点共有$\binom{180}{4}$个. 事件 A_k 中的样本点可以看作是"由 8 只次品中任取 k 只"且"从 172 只合格品中任取 $4-k$ 只"的结果，这种样本点（即 A_k 中的样本点）有$\binom{8}{k}\binom{172}{4-k}$个. 所以，

$$P(A_k)=\frac{\binom{8}{k}\binom{172}{4-k}}{\binom{180}{4}}.$$

对于 $k=0, 1, 2, 3, 4$，算出 $P(A_k)$ 列成下表：

表 1-4

k	0	1	2	3	4
$P(A_k)$	0.832	0.158	0.010	0.000	0.000

"次品超过 1 只"就是次品"恰有 2 只"或"恰有 3 只"或"恰有 4 只"，即 $B=A_2+A_3+A_4$，注意到 A_2, A_3, A_4 两两互不相容，利用可加性

$$P(B)=P(A_2)+P(A_3)+P(A_4)=0.010.$$

由此，还可以证明，当 180 只集成电路中含有不超过 8 只次品时，抽 4 只检查，出现 1 只以上次品的概率不超过 0.01，这个概率很小，在一次检查中这事件是实际上难得发生的. 如果在一次检查中，这事件真的发生了，我们便有理由怀疑这 180 只集成电路中不只 8 只是次品. 在产品抽样检查中就是根据这个思想来制定检查方案的.

这个例子的问题的一般情形，叙述如下：

设有 N 件产品，其中含有 M 件次品，从中任意抽取 n 件，"抽取的 n 件中恰好含有 k 件次品"（事件 A_k）的概率为

$$P(A_k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,\ 1,\ \cdots,\ n.$$

【例 2】 将 n 个不同的球任意放到 N 个盒中，每个盒子容纳球的个数不限. 试求 (1) 指定的 n 个盒子各放有一个球的概率 p_1；(2) 有 n 个盒子各放有一个球的概率 p_2.

解：这个随机试验是第一节例 4 的一般情形.

将 n 个球放入 N 个盒中，可以按下述步骤来完成：先将第一个球放入 N 个盒中的任意一个（有 N 种方法），然后将第二个球放入 N 个盒中的任意一个（有 N 种方法），…，最后将第 n 个球放入 N 个盒中的任意一个（有 N 种放法），所以，将 n 个球放入 N 个盒中有 N^n 种不同的放法. 每一种放法是一样本点，即样本点有 N^n 个. 由于放法的任意性，各样本点的出现是等可能的.

指定的 n 个盒子各放有一球，这种放法就是将 n 个球在 n 个指定的盒中的排列，有 $n!$ 种，所以

$$p_1=\frac{n!}{N^n}.$$

任何 n 个盒子各放有一个球，这种放法可按下述步骤来完成：先选出 n 个盒子（有 $\binom{N}{n}$ 种方法），然后将 n 个球分别放入 n 个选出的盒子中（有 $n!$ 种方法）. 所以，任何 n 个盒子各放有一个球的放法有 $\binom{N}{n}n!$ 种，从而，当 $N\geqslant n$ 时，得到

$$p_2=\frac{\binom{N}{n}n!}{N^n}=\frac{N!}{N^n(N-n)!},$$

而当 $N<n$ 时，$p_2=0$.

【例 3】 将 n 个相同的球任意放到 $N(N\geqslant n)$ 个盒中，每个盒子容纳球的个数不限. 如果各种不同的放法是等可能的，试求(1)指定的 n 个盒中各放有一球的概率 p_1，(2) 有 n 个盒子各放有一球的概率 p_2.

解：这个随机试验与前例不同，以三个球三个盒为例：在前例中 |ab|c| |，|ac|b| | 是两种不同的放法，在本例中，由于球相同，上述两个放法将不加分辨而算作一种放法. 于是不同的放法可图示如下：

|000| | |，|00|0| |，|00| |0|，|0|00| |，
|0|0|0|，|0| |00|，| |000| |，| |00|0|，
| |0|00|，| | |000|.

为了计算这种放法的个数，我们将三个盒连放在一起，用两个公共的墙"|"隔开. 每一种放法都对应着三个球 0, 0, 0 和两个（公共）墙 |、| 的一个排列. 这个对应是一一对应，所以，只需算出三个球 0, 0, 0 和两个墙 |、| 排列的个数. 每一个排列，有五个位置，每选出三个位置放球就得到一个这样的排列，所以，共有

$$\binom{5}{3}=\frac{5\times4\times3}{1\times2\times3}=10$$

种排法. 这样就求出了将3个相同的球任意放到3个盒中有$\binom{5}{3}$种放法.

一般, 将n个相同的球任意放入N个盒中(每盒容纳球的个数不限)有$\binom{n+N-1}{n}$种放法(求法与前面的特殊情形一样), 根据题设每种放法出现的可能性相同, 是古典概型.

指定的n个盒中各放有一个球的放法只有1种, 故得

$$p_1=\frac{1}{\binom{n+N-1}{n}}.$$

任何n个盒子各放有一个球有$\binom{N}{n}$种放法, 故

$$p_2=\frac{\binom{N}{n}}{\binom{n+N-1}{n}}=\frac{N!(N-1)!}{(N-n)!(n-N-1)!}.$$

【例4】 将n个相同的球任意放入$N(N\geqslant n)$个盒中, 每盒至多容纳1球, 若不同的放法是等可能的, 试求 (1) 指定n个盒子各放有一个球的概率p_1, (2) 有n个盒子各放有一个球的概率p_2.

解: 在这个随机试验中, 每一种放法就相当于从N个盒子中取出n个, 共有$\binom{N}{n}$种. 按题设它们的出现是等可能的.

指定的n个盒子各放有一球有一种放法, 故

$$p_1=\frac{1}{\binom{N}{n}}=\frac{n!(N-n)!}{N!}.$$

任何 n 个盒子各放有一球的放法有$\binom{N}{n}$种，故 $p_2=1$.

以上三例，如果将球解释为质点，盒子解释为空间中的小区域，就成为统计物理中的模型. 例 2 中假定质点可分辨，小区域可容纳任意个质点，是马克斯威-波尔茨曼 (Maxwell-Boltzmann) 提出的模型. 例 3 中所考虑的质点是不可分辨的，是波色-爱因斯坦(Bose-Einstein)提出的模型，适用于所谓波色子的基本粒子. 例 4 中考虑的是不可分辨的质点且假定小区域只能容纳一个质点，是费米-狄拉克 (Fermi-Dirac) 提出的模型，适用于所谓费米子的基本粒子.

由这些例子我们看到，一个随机试验中的样本点是什么，能否看作等可能的，需要根据对象的实际情形而定，这是必需注意的.

【例 5】 把 n 根棍子都折断成一长一短的两小段，把这 $2n$ 个小段任意配成 n 对，接成新的棍子，试求：

(1) 这些小段都连接成原来样子(A)的概率，

(2) 所有长的小段都与短的小段配对(B)的概率.

这个问题的背景是，当细胞被有害的放射线照射的时候，一些染色体就会破裂，这可以比作本例中的棍子，用“长的”小段表示含有中心体的那一部分，如果两个“长的”部分相结合或两个“短”的部分相结合，那么此细胞就会死亡.

解：把 $2n$ 个东西，分成两个一组，可以这样分：先从这 $2n$ 个东西中任取两个作为第 1 组，再从剩下的 $2n-2$ 个东西中任取两个作为第 2 组，……，最后剩下的两个作为第 n 组. 第 1 组有$\binom{2n}{2}$种取法，第 2 组有$\binom{2n-2}{2}$种取法，……，所以把 n 组都取出来共有

$$\binom{2n}{2}\binom{2n-2}{2}\cdots\binom{4}{2}\binom{2}{2}$$

种取法．但是只把组的次序重排，不动组内的东西，并不改变把 $2n$ 个东西分成两个一组的分法．所以，上述每 $n!$ 个取法对应于同一种分法．这样，在把 $2n$ 个小段棍子任意配成 n 对这一随机试验中，等可能的样本点的总数是

$$\frac{\binom{2n}{2}\binom{2n-2}{2}\cdots\binom{4}{2}\binom{2}{2}}{n!}=\frac{(2n)!}{2^n\cdot n!}.$$

(1) 只有一个样本点是这些小段都连接成原来的样子．所以

$$P(A)=1\Big/\frac{(2n)!}{2^n\cdot n!}=\frac{2^n\cdot n!}{(2n)!}.$$

(2) 先把"长"段按一个固定次序排好，如（长$_1$，长$_2$，…，长$_n$），然后让"短"段的任意一个全排列如（短$_{i_1}$，短$_{i_2}$，…，短$_{i_n}$）和"长"段配成

$$\begin{pmatrix}长_1 & 长_2 & \cdots & 长_n\\ 短_{i_1} & 短_{i_2} & \cdots & 短_{i_n}\end{pmatrix},$$

n 个"短"段的全排列数是 $n!$．事件 B 就是由这样的 $n!$ 个样本点组成．所以

$$P(B)=n!\Big/\frac{(2n)!}{2^n\cdot n!}=\frac{2^n(n!)^2}{(2n)!}.$$

在本节结束时，我们指出，本节建立古典概型的方法可以推广，得到下面的结论：

1．设样本空间 Ω 由有限个样本点组成，$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，$\mathscr{F}$ 是 Ω 的一切子集组成的事件域；又设 p_i 是一组给定的实数，满足

$$0\leqslant p_i\leqslant 1,\quad i=1,\cdots,n,\quad \sum_{i=1}^{n}p_i=1,\tag{1'}$$

则在 $\mathscr{F}$ 上存在一个概率 P，满足

$$P(\{\omega_i\})=p_i;\quad i=1,2,\cdots,n. \tag{2'}$$

这个概率的定义如下:

$$P(A)=p_{i_1}+\cdots+p_{i_m},\quad 当\ A=\{\omega_{i_1},\cdots,\omega_{i_m}\}; \tag{3'}$$
$$P(\emptyset)=0.$$

不难看出,当 $p_1=\cdots=p_n=\frac{1}{n}$ 时的特殊情形,就是古典概型. (3′)所定义的函数 P 确是 $\mathscr{F}$ 上的概率并满足 (2′),验证方法和古典概型完全类似,读者试完成之.

结论 1 还可以推广到一般离散样本空间的情形.

2. 设样本空间 Ω 由可数个样本点组成:

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n,\cdots\},$$

$\mathscr{F}$ 是 Ω 的一切子集组成的事件域,又设 p_i 是给定的实数序列,满足

$$\sum_{i=1}^{\infty}p_i=1,\quad 0\leqslant p_i\leqslant 1,\quad i=1,2,\cdots, \tag{1''}$$

则在 $\mathscr{F}$ 上存在一个概率 P, 满足

$$P(\{\omega_i\})=p_i,\quad i=1,2,\cdots. \tag{2''}$$

这个概率的定义为:

$$P(A)=p_{i_1}+p_{i_2}+\cdots+p_{i_m}+\cdots=\sum_{\omega_i\in A}p_i,$$
$$当\ A=\{\omega_{i_1},\omega_{i_2},\cdots,\omega_{i_m},\cdots\}; \tag{3''}$$
$$P(\emptyset)=0.$$

习 题 二

1. 一部四卷文集,按任意次序放到书架上,问各卷自左向右或自右向左的卷号顺序恰为 1, 2, 3, 4 的概率是多少?
2. 把 1, 2, 3, 4, 5 诸数各写在一张小纸片上,任取其三张排成自左向右的顺序,问如此所得三位数是偶数的概率是多少?
3. 一个电梯载着 7 个乘客,且止于 10 层楼上,试求不发生两位乘客在同一层上离开的概率.
4. 有 n 个人,设每人的生日是一年中的任何一天的概率为 1/365,试求此 n 个人生日互不相同的概率.
5. 盒中盛有 m 个白球,n 个黑球,从中任取 $a+b$ 个 $(0\leqslant a\leqslant m,\ 0\leqslant b$

$\leqslant n$)，试求所取的球恰含 a 个白球和 b 个黑球的概率？

6. 盒中有 m 个白球，n 个黑球，从中任意地接连取出 $k+1$ 个球（$k+1 \leqslant m+n$），每球被取出后不还原．试求最后取出的球是白球的概率．

7. 把 n 个不同的字母（其中有字母 A，B）排成一列，问恰好有两个字母排在 A，B 之间的概率是多少？如果排成一个圆圈，上面问题中的概率又是多少？

8. 从 0，1，2，…，9 中任取四个数，问能排成一个四位偶数的概率是多少？

9. 从 1，2，…，N 中随机地取出 n 个（每次取完以后都放回），问所取的 n 个数全不相同的概率是多少？

10. 将 n 个球投入 N 个盒中，每个球都以 $\dfrac{1}{N}$ 的概率进入每一盒中，试求指定的某盒是空盒的概率．

11. 将 n 个球投入 N 个盒中，设球不可辨（即只考虑每盒中的球数，而不区分是哪些球），每一种投法都是等概率的，试求：

    (1) 没有空盒的概率；

    (2) 正好有 m 个空盒的概率；

    (3) 指定 m 个盒中正好有 k 个球的概率．

12. 例 1 中若 180 只集成电路中的次品少于 8 只，试证抽验 4 只，出现多于 1 只次品的概率不超过 0.01．

第四节　几 何 概 型

向某一可度量的区域 G 内投一点，如果所投的点落在 G 中任意区域 g 内的可能性大小与 g 的度量成正比，而与 g 的位置和形状无关，则称这个随机试验为几何型随机试验，或称为几何概型．

上面所说的度量，指的是线段的长度，可求积平面区域的面积，可求积空间区域的体积，等等．

当一个投点的试验，如果根据其实际意义判断所投的点具有某种等可能性，或均匀性时，便可以看作是几何型的随机

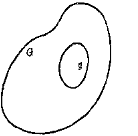


图 1-6

试验，这是因为对于这种试验设想重复进行了 n 次，所投的 n 个点应均匀地分布在 G 中各个位置，因而落在任意区域 g 内的频率就应与 g 的度量成正比，而与其位置和形状无关，所以符合几何概型的要求.

几何型随机试验的样本点可以用 G 中的点表示，因而样本空间就是 $\Omega=G$. 事件域 $\mathscr{F}$ 即取包含一切可度量区域的最小事件域.

现在来确定几何型随机试验中的概率 P. 从几何型随机试验的意义自然 P 应满足

$$P(A)=\lambda\times[A\text{ 的度量}].$$

由于 Ω 是必然事件,

$$1=P(\Omega)=\lambda\times[\Omega\text{ 的度量}],$$

$$\lambda=1/[\Omega\text{ 的度量}].$$

由此得到几何概型中概率的公式

$$P(A)=\frac{[A\text{ 的度量}]}{[\Omega\text{ 的度量}]}. \tag{1}$$

【例 1】(会面问题) 两个人约定于 12 点至 1 点在某地见面，先到者等 20 分钟后离去，试求两人能会面的概率.

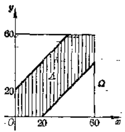


图 1-7

解：用 x, y 分别表示两人到达的时刻(12 时 x 分，12 时 y 分)，问题可以看作向平面区域

$$\Omega=\{(x,\ y)\,|\,0\leqslant x\leqslant 60,\ 0\leqslant y\leqslant 60\}$$

内投点．由于两人可分别"等可能的"在任何时刻到达，故可看作几何型随机试验．

"两人能会面"这事件就是

$$A=\{(x,\ y)\,|\,|x-y|\leqslant 20\},$$

$$P(A)=[A\ 的面积]/[\Omega\ 的面积]$$

$$=2000/3600=5/9.$$

【例 2】(投针问题)　平面上画有等距离为 $a(a>0)$ 的一些平行线，向平面任意投一长为 $l(l<a)$的针，试求针与平行线之一相交的概率．

解：设 x 是针的中点 M 到最近的平行线的距离，φ 是针与此线的交角．于是投针问题就相当于向平面区域

$$\Omega=\left\{(\varphi,\ x)\,|\,0\leqslant\varphi\leqslant\pi,\ 0\leqslant x\leqslant\frac{a}{2}\right\}$$

投点的几何型随机试验．

针与平行线相交的充分必要条件是$(\varphi,\ x)$满足

$$0\leqslant\frac{x}{\sin\varphi}\leqslant\frac{l}{2}\quad 即\quad 0\leqslant x\leqslant\frac{l}{2}\sin\varphi,$$

所以事件"针与平行线之一相交"就是

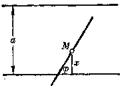


图 1-8

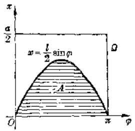


图 1-9

$$A=\left\{(\varphi,\ x)\ \middle|\ 0\leqslant x\leqslant \frac{l}{2}\sin\varphi,\ 0\leqslant\varphi\leqslant\pi\right\}.$$

它的概率

$$P(A)=\int_0^{\pi}\frac{l}{2}\sin\varphi\,d\varphi\Big/\frac{a}{2}\pi=\frac{2l}{\pi a}. \tag{2}$$

这个事件的概率与 π 有关，因此不少人想利用它来计算 π 的数值．这可以通过下述方法来实现：投针 n 次，计算针与平行线相交的次数 m，再以频率值 $\frac{m}{n}$ 作为概率 $P(A)$ 之值代入(2)，可得

$$\pi\approx\frac{2ln}{am}.$$

现在把这种试验的一些有关资料列于下表*)(其中把 a 折算为1)：

表 1-5

试验者	年份	针长	投掷次数	相交次数	π的试验值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan, C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

上面所用方法概括起来是：建立一个概率模型，使它与某些我们感兴趣的量(本例中是常数 π)有关，然后设计适当的随机试验，并通过这个试验的结果来确定这些量．随着电子计算机的发展，已按上述思路建立起一类新的方法，称为蒙特

*) 转引自复旦大学编概率论，第一册，第39页．原资料见 Gridgeman, N. T. Geometric probability and the number π. Scripta Mathematica, 25 (1960), 183～195.

卡洛(Monté-Carlo)方法.

从上面的试验记录中，我们也可以看到按几何概型所确定的概率与试验结果的一致性.

习 题 三

1. 两船欲靠同一码头,设两船独立地到达，而且各自到达时间在一昼夜间是等可能的. 如果此两船在码头的停留时间分别是 1 及 2 小时,试求一船要等待空出码头的概率.
2. 两人约好在某一地点会面. 在时间 T_1 到 T_2 之内到达. 假定他们在 T_1 至 T_2 之间的任何时刻到达都是等可能的. 试求其中一人必须等另外一人的时间不小于 t_0 的概率.
3. 在线段$(0, a)$上任意投三个点,试求由 0 至三点的三条线段能构成三角形的概率.
[提示: 看作投点(x, y, z), 其中 x, y, z 分别是 0 至三点的三条线段的长. 能构成三角形的充要条件是任意两线段长之和大于第三线段之长.]
4. 在一张打上方格的纸上投一枚直径为 1 的硬币, 方格要多大才能使硬币与线不相交的概率小于 0.01.
5. 从(0,1)中随机地取两个数,试求下列事件的概率: (1)"两数之和小于 1.5", (2)"两数之积小于 $\frac{1}{4}$", (3)"两数最大者小于 $\frac{1}{2}$", (4)"两数最小者小于 $\frac{1}{2}$".

第五节 概 率 的 性 质

由概率的定义可以导出它的一系列基本性质. 第二节引入概率的定义后,已经证明了两个性质,现在重述如下:

性质 1 $P(\emptyset)=0$.

性质 2 若 $A_1, \cdots, A_n$ 两两互不相容,则

$$P(A_1+\cdots+A_n)=P(A_1)+\cdots+P(A_n).$$

现在证明其它性质:

性质 3 $P(\overline{A})=1-P(A)$.

【证】 $A+\overline{A}=\Omega$, $A\overline{A}=\emptyset$, 利用性质 2,

$$1=P(\Omega)=P(A+\overline{A})=P(A)+P(\overline{A}).\quad]$$

性质 4 若 $A\subseteq B$, 则

$$P(B-A)=P(B)-P(A).$$

【证】 在 $A\subseteq B$ 时, $A+(B-A)=B$; 又

$$A\cap(B-A)=A\cap(B\cap\overline{A})=\emptyset,$$

故由性质 2,

$$P(B)=P(A)+P(B-A).\quad]$$

性质 5 若 $A\subseteq B$, 则 $P(A)\leqslant P(B)$.

【证】 由概率的非负性及性质 4,

$$P(B)-P(A)=P(B-A)\geqslant 0.\quad]$$

性质 6 $P(A+B)=P(A)+P(B)-P(AB)$.

【证】 $A+B=A+(B-AB)$, $A\cap(B-AB)=\emptyset$, $AB\subseteq B$, 利用性质 2 和性质 4,

$$P(A+B)=P(A)+P(B-AB)$$
$$=P(A)+P(B)-P(AB).\quad]$$

性质 7 若 $B_1\supseteq B_2\supseteq\cdots\supseteq B_n\supseteq\cdots$, 并且 $\bigcap_{i=1}^{\infty}B_i=\emptyset$, 则 $\lim_{n\to\infty}P(B_n)=0$.

【证】 令 $A_0=\bigcap_{i=1}^{\infty}B_i$, $A_i=B_i-B_{i+1}$, $i=1, 2, \cdots$, 则 A_1, A_2, $\cdots$ 两两互不相容, 并且对任何 n

$$B_n=A_0+A_n+A_{n+1}+\cdots=A_n+A_{n+1}+\cdots;$$

因而有 $$P(B_n)=\sum_{i=n}^{\infty}P(A_i);$$

特别$n=1$时，

$$P(B_1)=\sum_{i=1}^{\infty}P(A_i).$$

所以$\sum\limits_{i=n}^{\infty}P(A_i)$是收敛级数$\sum\limits_{i=1}^{\infty}P(A_i)$的余项，因而

$$\lim_{n\to\infty}P(B_n)=\lim_{n\to\infty}\sum_{i=n}^{\infty}P(A_i)=0. \quad \blacksquare$$

性质8(概率的上连续性) 若$A_1\supseteq A_2\supseteq\cdots\supseteq A_n\supseteq\cdots$，则

$$P\left(\bigcap_{n=1}^{\infty}A_n\right)=\lim_{n\to\infty}P(A_n).$$

【证】 令$B_n=A_n-\bigcap\limits_{i=1}^{\infty}A_i$，$n=1, 2, \cdots$，则$B_1\supseteq B_2\supseteq\cdots\supseteq B_n\supseteq\cdots$，并且$\bigcap\limits_{n=1}^{\infty}B_n=\emptyset$，由性质7，

$$\begin{aligned}\lim_{n\to\infty}P(A_n)&=\lim_{n\to\infty}P\left(\left(A_n-\bigcap_{i=1}^{\infty}A_i\right)+\bigcap_{i=1}^{\infty}A_i\right)\\&=\lim_{n\to\infty}P(B_n)+P\left(\bigcap_{i=1}^{\infty}A_i\right)\\&=P\left(\bigcap_{i=1}^{\infty}A_i\right). \quad \blacksquare\end{aligned}$$

性质9(概率的下连续性) 若$A_1\subseteq A_2\subseteq\cdots\subseteq A_n\subseteq\cdots$，则

$$P\left(\bigcup_{n=1}^{\infty}A_n\right)=\lim_{n\to\infty}P(A_n).$$

【证】 $\overline{A}_1\supseteq\overline{A}_2\supseteq\cdots\supseteq\overline{A}_n\supseteq\cdots$，由性质8，

$$P\left(\bigcap_{n=1}^{\infty}\overline{A}_n\right)=\lim_{n\to\infty}P(\overline{A}_n),$$

$$\begin{aligned}\lim_{n\to\infty}P(A_n)&=1-\lim_{n\to\infty}P(\overline{A}_n)=1-P\left(\overline{\bigcup_{n=1}^{\infty}A_n}\right)\\&=P\left(\bigcup_{n=1}^{\infty}A_n\right). \quad \blacksquare\end{aligned}$$

习 题 四

1. A、B 互不相容，$P(A)$、$P(B)$ 已知，试求 $P(A\cup B)$，$P(A\cap B)$，$P(A\cup\bar{B})$，$P(A\cap\bar{B})$，$P(\bar{A}\cap\bar{B})$。

2. 已知 $P(A)=p$，$P(B)=q$，$P(A\cup B)=r$，求 $P(A\bar{B})$ 及 $P(\bar{A}\bar{B})$。

3. 证明：$P(A\bar{B}+\bar{A}B)=P(A)+P(B)-2P(AB)$。

4. 证明：$P(A_1\cup\cdots\cup A_n)=1-P(\bar{A}_1\cdots\bar{A}_n)$。

5. 证明 $P(A_1\cup\cdots\cup A_n)=P(A_1)+P(\bar{A}_1A_2)+P(\bar{A}_1\bar{A}_2A_3)+\cdots+P(\bar{A}_1\cdots\bar{A}_{n-1}A_n)$。

6. (1) 证明

$$\begin{aligned}P(A_1\cup A_2\cup A_3)&=P(A_1)+P(A_2)+P(A_3)\\&\quad-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)\\&\quad+P(A_1A_2A_3);\end{aligned}$$

(2) 利用数学归纳法证明

$$\begin{aligned}&P(A_1\cup A_2\cup\cdots\cup A_n)\\&=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i_1<i_2\leqslant n}P(A_{i_1}A_{i_2})+\sum_{1\leqslant i_1<i_2<i_3\leqslant n}P(A_{i_1}A_{i_2}A_{i_3})-\cdots\\&\quad+(-1)^{k-1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}P(A_{i_1}A_{i_2}\cdots A_{i_k})+\cdots\\&\quad+(-1)^{n-1}P(A_1A_2\cdots A_n).\end{aligned}$$

7. 已知 $P(AB)=P(A)P(B)$，$C\supseteq AB$，$\bar{C}\supseteq\bar{A}\bar{B}$，证明

$$P(AC)\geqslant P(A)P(C).$$

8. (1) 已知：A_1 和 A_2 同时发生则 A 发生，试证

$$P(A)\geqslant P(A_1)+P(A_2)-1;$$

(2) 已知 $A_1A_2A_3\subseteq A$，试证 $P(A)\geqslant P(A_1)+P(A_2)+P(A_3)-2$。

9. 证明

(1) 若 $P(A_i)=0(i=1,\ 2,\ \cdots)$，则 $P\left(\bigcup\limits_{i=1}^{n}A_i\right)=0$；

(2) 若 $P(A_i)=1(i=1,\ 2,\ \cdots)$，则 $P\left(\bigcap\limits_{i=1}^{n}A_i\right)=1$。

10. 将 n 个球投入 N 个盒中，每个球都以 $\dfrac{1}{N}$ 的概率进每一盒中，试求至少有一球不进入指定盒的概率。

11. 某人写了 n 个通讯地址，在每个信封里各封入一封信，然后在每个信封上任意抄写一个前面所说的地址．问至少有一封信碰对了地址的概率是多少？

第六节　条件概率

在实际问题中，除了要知道事件 A 的概率 $P(A)$ 外，有时还需要知道在"事件 B 已发生"的条件下，事件 A 的概率 $P(A|B)$．由于增加了新的条件"事件 B 已发生"，所以 $P(A|B)$ 一般说来与 $P(A)$ 不同，而称 $P(A|B)$ 为条件概率．

现在利用频率来分析一下应该怎样定义 $P(A|B)$．

设 A，B 是某随机试验中的两个事件，进行了 n 次试验，其中事件 B，AB 各发生 m_B，m_{AB} 次．这时，"事件 B 已发生"条件下的试验有 m_B 次，其中事件 A 发生 m_{AB} 次．所以，

$$\text{"事件 } B \text{ 已发生"条件下 } A \text{ 发生的频率} = \frac{m_{AB}}{m_B}$$

$$= \frac{m_{AB}/n}{m_B/n} = \frac{AB \text{ 发生的频率}}{B \text{ 发生的频率}}.$$

由于频率与概率的相近，因此我们

定义　设 A，B 是两个事件，$P(B)>0$，称

$$P(A|B) = \frac{P(AB)}{P(B)}$$

为在事件 B 已发生的条件下，事件 A 的条件概率．

在古典概型中，设 Ω 由 n 个样本点组成，若事件 B 包含 m 个样本点 $(m>0)$，AB 包含 k 个样本点，则

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{k/n}{m/n} = \frac{k}{m}.$$

在几何概型中，

$$P(A|B)=\frac{[AB\text{ 的度量}]}{[B\text{ 的度量}]}.$$

由条件概率的定义，立即可以得到

乘法公式 $P(AB)=P(A|B)P(B)$，当 $P(B)>0$.

这个公式在概率计算中有重要作用.

【例 1】 罐中有两个白球，三个黑球，从中依次取出两个，试求取出的两个球都是白球(事件 A)的概率.

解：用 A_1 表示"第一次取得白球"，A_2 表示"第二次取得白球"，则 $A=A_1A_2$.

在 A_1 已发生时，罐中剩有一个白球三个黑球，所以，

$$P(A_2|A_1)=1/4.$$

于是利用乘法公式，得到

$$P(A)=P(A_1A_2)=P(A_1)P(A_2|A_1)=\frac{2}{5}\times\frac{1}{4}=\frac{1}{10}.$$

【例 2】 有三个形状相同的罐，在第一个罐中有两个白球和一个黑球；在第二个罐中有三个白球和一个黑球；在第三个罐中有两个白球和两个黑球。某人任取一罐，从中任取一球，试求取得白球的概率.

解：用 A 表示"取得白球"，H_i 表示"球取自第 i 罐"($i=1, 2, 3$)，便有

$$A=AH_1+AH_2+AH_3.$$

利用可加性及乘法公式，

$$\begin{aligned}P(A)&=P(AH_1)+P(AH_2)+P(AH_3)\\&=P(H_1)P(A|H_1)+P(H_2)P(A|H_2)\\&\quad+P(H_3)P(A|H_3)\\&=\frac{1}{3}\times\frac{2}{3}+\frac{1}{3}\times\frac{3}{4}+\frac{1}{3}\times\frac{1}{2}=\frac{23}{36}.\end{aligned}$$

上例中确定 $P(A)$ 的方法具有普遍意义，可以概括成

定理1（全概率公式） 若事件 $H_1, \cdots, H_n$ 两两互不相容，$P(H_i)>0\ (i=1, \cdots, n)$，并且 $A\subseteq\bigcup_{i=1}^{n}H_i$，则

$$P(A)=\sum_{i=1}^{n}P(H_i)P(A|H_i).$$

全概率公式的想法是将确定 A 的概率问题分解为确定 $AH_1, \cdots, AH_n$ 的概率.

它的证明方法实质上就是例2的解法的一般化.

定理2（贝叶斯(Bayes)公式） 若事件 $H_1, \cdots, H_n$ 两两互不相容，$P(H_i)>0(i=1,\cdots, n)$，且 $A\subseteq\bigcup_{i=1}^{n}H_i$，$P(A)>0$，则

$$P(H_k|A)=\frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n}P(H_k)P(A|H_i)}\quad (k=1, \cdots, n).$$

【证】 由乘法公式及全概率公式，对任一 $H_k(1\leqslant k\leqslant n)$，

$$P(H_k|A)=\frac{P(AH_k)}{P(A)}=\frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n}P(H_i)P(A|H_i)}.\quad \blacksquare$$

【例3】 根据以往的临床记录，某种诊断癌症试验具有如下的效果：以 A 表示"试验反应为阳性"，C 表示"被诊断者患有癌症"，

$$P(A|C)=0.95;\ P(\overline{A}|\overline{C})=0.95.$$

现在对一大批人进行癌症普查，设被试验的人中患有癌症的概率为0.005，即 $P(C)=0.005$，试求 $P(C|A)$——一人试验反应为阳性的情况下，此人确患癌症的概率.

解：利用贝叶斯公式，

$$P(C|A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\bar{C})P(A|\bar{C})}$$
$$=\frac{0.005\times 0.95}{0.005\times 0.95+0.995\times 0.05}=0.087.$$

结果表明，虽然 $P(A|C)$，$P(\bar{A}|\bar{C})$ 比较大，但试验呈阳性的人确患癌症的可能性还是不大.

【例 4】 无线电通讯中，由于随机干扰，当发出信号为"·"时，收到信号为"·"、"不清"和"-"的条件概率分别为 0.7、0.2、0.1. 当发出的信号为"-"时，收到信号为"-"、"不清"和"·"的条件概率分别是 0.9、0.1、0. 如果发报过程中"·"和"-"出现的概率分别是 0.6 和 0.4，当收到讯号"不清"时，原发出的信号是什么？试加以推测.

解：用 H_1 表示"原发出信号为"·"，

H_2 表示"原发出信号为"-"，

A 表示"收到信号"不清".

今计算 $P(H_1|A)$ 及 $P(H_2|A)$ 并加以比较：

$$P(H_1|A)=\frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1)+P(H_2)P(A|H_2)}$$
$$=\frac{0.6\times 0.2}{0.6\times 0.2+0.4\times 0.1}=0.75,$$
$$P(H_2|A)=\frac{P(H_2)P(A|H_2)}{P(H_1)P(A|H_1)+P(H_2)P(A|H_2)}$$
$$=\frac{0.4\times 0.1}{0.6\times 0.2+0.4\times 0.1}=0.25.$$

由此可见，当收到信号不清时，原发出信号为"·"比之为"-"有较大的概率. 因此通常可以推断原发生信号为"·".

习 题 五

1. 某种集成电路使用到 2000 小时还能正常工作的概率是 0.94，使

用到 3000 小时还能正常工作的概率是 0.87，问已经工作了 2000 小时的集成电路能继续工作到 3000 小时的概率是多少？

2．某射击小组共有 20 名射手，其中一级射手 4 人，二级射手 8 人，三级射手 7 人，四级射手 1 人，一、二、三、四级射手能通过选拔进入比赛的概率分别为 0.9、0.7、0.5、0.2，求任取一位射手能通过选拔进入比赛的概率．

3．从装有 3 个白球与 2 个黑球的盒子中取出两个球移放到装有 4 个白球与 4 个黑球的盒子中，试计算从第二个盒子中取出一球是白球的概率．

4．甲盒中装有 5 个白球与 5 个黑球，从中取出 5 个球放入空盒乙中，再从乙盒中取出 3 个球放入空盒丙中，最后从丙盒中取出一个球是白球，试求第一次取出的 5 个球全是白球的概率．

5．一批螺丝钉，由甲、乙、丙三机床生产，各占总量的 25%、35%、40%，各机床产品的废品率分别为 5%、4%、2%．从这批螺丝钉中随机取出一只经检查是废品，问这件废品是甲、乙、丙生产的概率分别等于多少？

6．某工厂生产的产品合格率是 0.96．为了保证出厂产品质量，需要进行检查，但直接检查带有破坏性，因此提出一种非破坏性的但不完全准确的简化检查方法．经试验知一个合格品经简化检查而获准出厂的概率是 0.98，而一个废品经简化检查而获准出厂的概率是 0.05．问使用这种简化检查方法，获得出厂许可的产品是合格品的概率及未获出厂许可的产品是废品的概率各是多少？

第七节　事件的独立性

定义　称事件 A 对事件 B **独立**，如果

$$P(A|B)=P(A) \quad \text{或} \quad P(B)=0.$$

A 对 B 独立的含义是事件 A 的概率不受附加条件"事件 B 已发生"的影响．

【例 1】　从含有 a 个白球和 b 个黑球的口袋中依次取出

两个球，考虑事件 B："第一次取到白球"，A："第二次取到白球"，C："第二次取到黑球"．

（i）不放回抽样情形，即假定每次取出球后不放回．

样本点是从 $a+b$ 个球中任取两个的不重复排列，共有 $(a+b)(a+b-1)$ 个．

B 中样本点是第一次从 a 个白球任取一个，然后再从其余的 $a+b-1$ 个球任取一个排成，有 $a(a+b-1)$ 个．

A 中样本点是第二次从 a 个白球任取一个与第一次从另外的 $a+b-1$ 个球任取一个排成，有 $a(a+b-1)$ 个．

类似地，C 有 $b(a+b-1)$ 个样本点．AB 有 $a(a-1)$ 个样本点，CB 有 ab 个样本点．

于是按古典概型求得

$$P(A)=P(B)=\frac{a}{a+b},\quad P(C)=\frac{b}{a+b},$$

$$P(AB)=\frac{a(a-1)}{(a+b)(a+b-1)},$$

$$P(CB)=\frac{ab}{(a+b)(a+b-1)},$$

$$P(A|B)=P(AB)/P(B)=\frac{a-1}{a+b-1}\neq P(A),$$

$$P(C|B)=P(CB)/P(B)=\frac{b}{a+b-1}\neq P(C).$$

在不放回抽样时，第一次取球的情况对第二次取球的情况有影响，所以，$P(A|B)\neq P(A)$，$P(C|B)\neq P(C)$．

（ii）放回抽样情形，即假定每次取出球后放回．

在这种情形下，样本点是从 $a+b$ 个球中任取两个的可重复排列，共有 $(a+b)^2$ 个．

B 中样本点是第一次从 a 个白球任取一个与第二次从

$a+b$ 个球任取一个排成，有 $a(a+b)$ 个.

A 中样本点是第二次从 a 个白球任取一个与第一次从 $a+b$ 个球任取一个排成，有 $a(a+b)$ 个.

类似地，C 有 $b(a+b)$ 个样本点，AB 有 a^2 个样本点，CB 有 ab 个样本点.

于是按古典概型求得

$$P(A)=P(B)=\frac{a}{a+b},\quad P(C)=\frac{b}{a+b},$$

$$P(AB)=\frac{a^2}{(a+b)^2},\ P(CB)=\frac{ab}{(a+b)^2},$$

$$P(A|B)=\frac{a}{a+b}=P(A),\ P(C|B)=\frac{b}{a+b}=P(C).$$

由于放还抽样，第一次取球的情况对第二次取球的情况没有影响，所以有上面得到的结论.

定理 1 事件 A 对 B 独立的充要条件是

$$P(AB)=P(A)P(B).$$

【证】 若 A 对 B 独立，则当 $P(B)>0$ 时，

$$P(AB)=P(A|B)P(B)=P(A)P(B);$$

当 $P(B)=0$ 时，$0\leqslant P(AB)\leqslant P(B)=0$，也有

$$P(AB)=0=P(A)P(B).$$

反之，若 $P(AB)=P(A)P(B)$，则当 $P(B)>0$ 时，

$$P(A)=\frac{P(AB)}{P(B)}=P(A|B),$$

由定义知 A 对 B 独立；当 $P(B)=0$ 时，按定义也有 A 对 B 独立. 】

推论 1 A 对 B 独立当且仅当 B 对 A 独立.

【证】 A 对 B 独立 $\Leftrightarrow P(AB)=P(A)P(B)\Leftrightarrow B$ 对 A 独立. 】

由此可见事件 A 和 B 之间的独立性是相互的，以后就说 A, B 独立.

推论2 若 A、B 独立，则 A、$\bar{B}$ 独立，$\bar{A}$、B 独立，$\bar{A}$、$\bar{B}$ 独立.

【证】 这里只证 A、$\bar{B}$ 独立，其余部分留作习题. 由定理1知只需证明 $P(A\bar{B})=P(A)P(\bar{B})$.

$$P(A\bar{B})=P(A-AB)=P(A)-P(AB).$$

因 A, B 独立，$P(AB)=P(A)P(B)$，所以

$$P(A\bar{B})=P(A)-P(A)P(B)$$
$$=P(A)(1-P(B))=P(A)P(\bar{B}).$$ 】

【例2】 考虑向正方形区域 $ABCD$ 内任意投一点的随机试验. 令 E, F, G, H 是正方形各边的中点(参看图1-10).

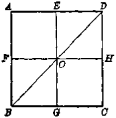


图 1-10

M_1 表示事件:"点子落在矩形 $AFHD$ 内",

M_2 表示事件:"点子落在矩形 $ABGE$ 内",

M_3 表示事件:"点子落在三角形 ABD 内".

试问 M_1, M_2, M_3 两两之间是否独立?

解：这是一个几何型随机试验，可以得出有关事件的概率(不妨设 $ABCD$ 的面积为1).

$$P(M_1)=P(M_2)=P(M_3)=1/2,$$

$$P(M_1M_2)=\frac{1}{4} \quad (\text{矩形 } AFOE \text{ 的面积}),$$

$$P(M_1M_3)=\frac{3}{8} \quad (\text{梯形 } AFOD \text{ 的面积}),$$

$$P(M_2M_3)=\frac{3}{8}(\text{梯形 }ABOE\text{ 的面积}).$$

利用定理1立即看出 M_1, M_2 独立；M_1,M_3 不独立；M_2,M_3 不独立.

在实际问题中，往往对两个事件 A、B，按其实际意义来看彼此是否有影响，从而判断其是否独立．若两个事件独立，就可以由两个事件概率的乘积来确定事件乘积的概率.

【例3】 甲乙两人对同一目标射击，甲击中目标的概率是0.9，乙击中目标的概率是0.8，甲乙两人各射击一次，试求击中目标的概率.

解："击中目标"(事件 C)就是"甲击中目标"(事件 A)或"乙击中目标"(事件 B)，即 $C=A+B$．于是

$$P(C)=P(A+B)=P(A)+P(B)-P(AB).$$

因为甲是否击中目标不受乙是否击中目标的影响，所以 A, B 独立，因而有

$$P(AB)=P(A)P(B)=0.9\times0.8=0.72,$$

$$P(C)=0.8+0.9-0.72=0.98.$$

事件独立的概念可以推广到任意有限多个事件上.

定义 n 个事件 $A_1, A_2, \cdots, A_n$ 称为相互独立，如果对于其中任意 m 个事件 $A_{i_1}, \cdots, A_{i_m}$ 有

$$P(A_{i_1}A_{i_2}\cdots A_{i_m})=P(A_{i_1})\cdots P(A_{i_m}),$$

$$1\leqslant i_1<i_2<\cdots<i_m\leqslant n,\quad 1\leqslant m\leqslant n.$$

上式代表 2^n-n-1 个等式：

$$P(A_{i_1}A_{i_2})=P(A_{i_1})P(A_{i_2}),\qquad \binom{n}{2}\text{个};$$

$$P(A_{i_1}A_{i_2}A_{i_3})=P(A_{i_1})P(A_{i_2})P(A_{i_3}),\qquad \binom{n}{3}\text{个};$$

……………… ……

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n),\qquad \binom{n}{n}\text{个}.$$

以三个事件为例，A_1, A_2, A_3 相互独立，要求下列四个等式成立：

$$P(A_1A_2)=P(A_1)P(A_2),\ P(A_1A_3)=P(A_1)P(A_3),$$
$$P(A_2A_3)=P(A_2)P(A_3),$$
$$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3).$$

需要注意的是若事件 $A_1, A_2, \cdots, A_n$ 两两独立（即其中每两个都独立），并不能导出它们相互独立. 我们可以举出下面的例子.

【例 4】 设 $\Omega=\{\omega_1, \omega_2, \omega_3, \omega_4\}$ 是等可能的，即

$$P(\{\omega_i\})=\frac{1}{4}\quad(i=1, 2, 3, 4).$$

令 $\quad A=\{\omega_1, \omega_2\},\ B=\{\omega_1, \omega_3\},\ C=\{\omega_1, \omega_4\}.$

有 $\quad P(A)=P(B)=P(C)=\frac{1}{2},$

$$P(AB)=P(AC)=P(BC)=P(\{\omega_1\})=\frac{1}{4}$$
$$=P(A)P(B)=P(A)P(C)=P(B)P(C),$$

因而事件 A, B, C 两两独立. 然而它们却不相互独立，因为

$$P(ABC)=P(\{\omega_1\})=\frac{1}{4}\neq\frac{1}{8}=P(A)P(B)P(C).$$

由相互独立的定义直接可以看出下述事实：

推论 3 如果 $A_1, A_2, \cdots, A_n$ 相互独立，那末其中任意 m 个事件也相互独立.

推论 4 如果事件 $A_1, A_2, \cdots, A_n$ 相互独立，则将其中任意个事件换成对立事件仍然相互独立，即有 $\bar{A}_1, A_2, \cdots, A_n$ 相互独立，$\bar{A}_1, \bar{A}_2, A_3, \cdots, A_n$ 相互独立，$\cdots$，$\bar{A}_1, \bar{A}_2, \cdots, \bar{A}_{n-1}, A_n$ 相互独立，$\bar{A}_1, \bar{A}_2, \cdots, \bar{A}_n$ 相互独立，等等. 一般地，

$$A_1, \cdots, \overline{A}_{i_1}, \cdots, \overline{A}_{i_k}, \cdots, \overline{A}_{i_m}, \cdots, A_n$$

相互独立.

【证】 我们只需证明将任意一个事件换成对立事件仍然相互独立，因为反复运用这一结论即可得出将任意个事件换成对立事件的结论.

又不失一般性，我们只需证明

$$\overline{A}_1, A_2, \cdots, A_n$$

相互独立. 为此我们只需验证

$$P(\overline{A}_1 A_{i_1} \cdots A_{i_m}) = P(\overline{A}) P(A_{i_1}) \cdots P(A_{i_m}).$$

事实上，

$$\begin{aligned} P(\overline{A}_1 A_{i_1} \cdots A_{i_m}) &= P(A_{i_1} \cdots A_{i_m}) - P(A_1 A_{i_1} \cdots A_{i_m}) \\ &= P(A_{i_1}) \cdots P(A_{i_m}) - P(A_1) P(A_{i_1}) \cdots P(A_{i_m}) \\ &= (1 - P(A_1)) P(A_{i_1}) \cdots P(A_{i_m}) \\ &= P(\overline{A}_1) P(A_{i_1}) \cdots P(A_{i_m}). \quad \blacksquare \end{aligned}$$

独立是概率论的一个重要概念，下面看看它在概率计算中的应用.

【例 5】 对同一目标进行三次射击，第一、二、三次射击的命中概率分别是 0.4、0.5、0.7，试求在这三次射击中恰有一次命中的概率.

解："恰有一次击中"(用 A 表示)是指第一次命中(用 A_1 表示)而第二、第三次都落空了，或第二次命中 (用 A_2 表示) 而第一、第三次都落空了，或第三次命中 (用 A_3 表示) 而第一、第二次都落空了，所以

$$A = A_1 \overline{A}_2 \overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1 \overline{A}_2 A_3.$$

由于三次射击是独立的，A_1, A_2, A_3 独立，利用可加性和独立性

$$
\begin{aligned}
P(A) &= P(A_1\bar{A}_2\bar{A}_3)+P(\bar{A}_1A_2\bar{A}_3)+P(\bar{A}_1\bar{A}_2A_3)\\
&= P(A_1)P(\bar{A}_2)P(\bar{A}_3)+P(\bar{A}_1)P(A_2)P(\bar{A}_3)\\
&\quad +P(\bar{A}_1)P(\bar{A}_2)P(A_3)\\
&= 0.4\times 0.5\times 0.3+0.6\times 0.5\times 0.3\\
&\quad +0.6\times 0.5\times 0.7=0.36.
\end{aligned}
$$

【例 6】 一个元件能正常工作的概率叫做这个元件的可靠性．由元件组成的系统能正常工作的概率叫做系统的可靠性．设构成系统的每个元件的可靠性均为 $r(0<r<1)$，且各元件能否正常工作是相互独立的．设以 $2n(n>1)$ 个元件按照下面图示的两种联接方式构成两个系统，试求它们的可靠性，并比较两个系统可靠性的大小．

系统 I：

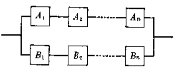


图 1-11

系统 II：

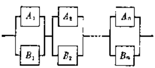


图 1-12

解：先讨论系统 I．它有两条通路．每条通路正常工作，当且仅当该通路上各个元件都能正常工作．故这两条通路的可靠性各为

$$P(A)=P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)=r^n,$$

$$P(B)=P(B_1B_2\cdots B_n)=P(B_1)P(B_2)\cdots P(B_n)=r^n.$$

(这里 $A=A_1A_2\cdots A_n$，$B=B_1B_2\cdots B_n$，而 A_i，B_i 分别表示事件"相应的元件能正常工作")．由于系统 I 是由两条通路并联而成的，它能正常工作，当且仅当至少有一条能正常工作．因此系统 I 的可靠性为

$$R_1=P(A+B)=1-P(\overline{A+B})=1-P(\bar{A}\bar{B})$$
$$=1-P(\bar{A})P(\bar{B})=1-(1-r^n)^2$$
$$=r^n(2-r^n).$$

对于系统 II, 每对并联元件的可靠性为

$$P(C_i)=P(A_i+B_i)=1-P(\bar{A}_i)P(\bar{B}_i)$$
$$=1-(1-r)^2=r(2-r),$$

这里 $C_i=A_i+B_i$, 系统是由各对并联元件串联而成的, 所以它的可靠性为

$$R_2=P(C_1C_2\cdots C_n)=P(C_1)P(C_2)\cdots P(C_n)$$
$$=r^n(2-r)^n.$$

从所得的结果看出, 系统 I (表示增加通路) 的可靠性和系统 II(表示采用附加元件)的可靠性都比单条通路大.

上面两个系统都是由 $2n$ 个可靠性相同的元件构成的,但 $R_2>R_1$(可用数学归纳法证明),因此系统 II 比系统 I 的可靠性大.

习 题 六

1. 证明(1) 若 $P(A)=1$, 则 A 与任意事件 B 独立;
   (2) 若 A 与任意事件 B 独立,则 $P(A)=0$ 或 1.
2. 若 $P(A)>0$, $P(B)>0$, 则 A, B 独立与 A, B 互不相容不能同时成立.
3. 已知三个事件两两独立,但不能三个同时发生, 假设它们有同一概率 x, 试决定 x 的最大可能值.
4. A, B 独立, $P(A\cup B)=0.6$, $P(A)=0.4$, $P(B)=?$
5. 证明事件 A_1, A_2, $\cdots$, A_n 相互独立的充分与必要条件是下面 2^n 个等式成立:
   $$P(C_1C_2\cdots C_n)=P(C_1)P(C_2)\cdots P(C_n),$$
   其中 $C_i=A_i$ 或 $C_i=\bar{A}_i$, $i=1, 2, \cdots, n$.

第八节　n 重贝努里试验

将某一试验（如投硬币试验）独立地重复进行 n 次（投 n 次硬币）是概率论中广泛讨论的一类随机试验．只考虑两个可能结果的试验称为贝努里（Bernoulli）试验（如投一个硬币观察出现正反面的试验，射击一次观察是否命中目标的试验等）．将一贝努里试验独立地重复 n 次的随机试验称为 n 重贝努里试验．在本节中我们讨论如何确定 n 重贝努里试验中事件的概率．

先结合一个例子来讨论．

【例 1】　某门炮向目标射击一次，命中的概率为 p，问这门炮向目标独立射击四次，恰好命中两次（事件 B）的概率是多少？

解：这个问题中的随机试验是"向目标独立射击 4 次（即 4 次独立试验）"，它的样本空间 Ω 由下列 $2^4=16$ 个样本点组成：

(0, 0, 0, 0),

(0, 0, 0, 1), (0, 0, 1, 0), (0, 1, 0, 0), (1, 0, 0, 0),

(0, 0, 1, 1), (0, 1, 0, 1), (0, 1, 1, 0), (1, 0, 0, 1);

(1, 0, 1, 0), (1, 1, 0, 0),

(0, 1, 1, 1), (1, 0, 1, 1), (1, 1, 0, 1), (1, 1, 1, 0);

(1, 1, 1, 1).

其中"0"代表"未命中"，"1"代表"命中"，(1, 0, 0, 1) 表示"第一、第四次命中，第二、第三次未命中"等等．由于样本点只有有限个，Ω 的每一子集都作为事件（即取事件域 $\mathscr{F}$ 包含 Ω 的一切子集）．

设 A_i 表示事件"第 i 次命中" ($i=1, 2, 3, 4$)，则各基本事件分别可表示为：

$\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4$,

$\overline{A}_1\overline{A}_2\overline{A}_3A_4$, $\overline{A}_1\overline{A}_2A_3\overline{A}_4$, $\overline{A}_1A_2\overline{A}_3\overline{A}_4$, $A_1\overline{A}_2\overline{A}_3\overline{A}_4$,

$\overline{A}_1\overline{A}_2A_3A_4$, $\overline{A}_1A_2\overline{A}_3A_4$, $\overline{A}_1A_2A_3\overline{A}_4$, $A_1\overline{A}_2\overline{A}_3A_4$,

$A_1\overline{A}_2A_3\overline{A}_4$, $A_1A_2\overline{A}_3\overline{A}_4$,

$\overline{A}_1A_2A_3A_4$, $A_1\overline{A}_2A_3A_4$, $A_1A_2\overline{A}_3A_4$, $A_1A_2A_3\overline{A}_4$,

$A_1A_2A_3A_4$.

根据问题的实际意义，这个随机试验中的概率 P 满足下述两个条件：

（i） $P(A_i)=p$ (因为已知在一次射击中命中的概率为 p)，$i=1, 2, 3, 4$.

（ii） A_1, A_2, A_3, A_4 独立 (因为试验是独立的，各次试验下的事件相互独立)。

于是，由概率的定义及条件（i），（ii），可以确定出各基本事件的概率，从而确定出所有事件的概率：

$$P(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4)=P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)P(\overline{A}_4)=q^4,$$

$$P(\overline{A}_1\overline{A}_2\overline{A}_3A_4)=P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)P(A_4)=pq^3.$$

类似地，

$$P(A_1\overline{A}_2\overline{A}_3\overline{A}_4)=P(\overline{A}_1A_2\overline{A}_3\overline{A}_4)=P(\overline{A}_1\overline{A}_2A_3\overline{A}_4)=pq^3,$$

$$P(\overline{A}_1\overline{A}_2A_3A_4)=P(\overline{A}_1A_2\overline{A}_3A_4)=\cdots$$
$$=P(A_1A_2\overline{A}_3\overline{A}_4)=p^2q^2,$$

$$P(\overline{A}_1A_2A_3A_4)=P(A_1\overline{A}_2A_3A_4)=P(A_1A_2\overline{A}_3A_4)$$
$$=P(A_1A_2A_3\overline{A}_4)=p^3q,$$

$$P(A_1A_2A_3A_4)=p^4,$$

其中 $q=1-p$.

$$\therefore \quad P(B)=P(\overline{A}_1\overline{A}_2A_3A_4)+P(\overline{A}_1A_2\overline{A}_3A_4)+P(\overline{A}_1A_2A_3\overline{A}_4)+P(A_1\overline{A}_2\overline{A}_3A_4)+P(A_1A_2\overline{A}_3\overline{A}_4)+P(A_1\overline{A}_2A_3\overline{A}_4) \quad \left(\binom{4}{2}\text{次}\right)$$

$$=\binom{4}{2}p^2q^2.$$

以上是一个4重贝努里试验的例子．利用完全类似的方法可以得出一般 n 重贝努里试验中的结论．

n 重贝努里试验是具有下述两个特点的随机试验：

(1) 试验的内容是重复 n 次某一试验，每次试验有两个可能的结果：A("成功")，$\overline{A}$("失败")且

$$P(A)=p, \quad P(\overline{A})=1-p=q,$$

与试验的序号 i 无关；

(2) n 次试验是独立进行的(即每次试验结果出现的概率不受其它各次试验结果的影响)，令 A_i 表示事件："第 i 次试验成功"，则 $A_1, \cdots, A_n$ 独立．

在 n 重贝努里试验中，事件 B_k："在 n 次试验中恰好有 k 次成功"的概率

$$P(B_k)=\binom{n}{k}p^kq^{n-k}, \quad k=0, 1, \cdots, n.$$

在 n 重贝努里试验中，n 次试验是同一试验的 n 次重复，如例1中，同一门炮独立地射击4次；不过也可以应用到 n 次不同的独立试验，只要每次试验中"成功"的概率相同，例如，4门炮独立地各向目标射击一次，如果每门炮命中的概率相同，也可以看作是4重贝努里试验．

【例2】 某小组有10台7.5千瓦的机床．如果每台机床的使用情况是互相独立的，且每台机床平均每小时开动12分钟，问全部机床用电超过48千瓦的可能性有多少？

解：在任一时刻，任何一台机床有"开动"、"停止"两种状态，用 A_i 表示"第 i 台机床开动"，则 $\overline{A}_i$ 表示："第 i 台机床停止"。由题意，$P(A_i)=\dfrac{12}{60}=0.2$，$P(\overline{A}_i)=0.8$。用 B_k 表示"恰有 k 台机床同时开动"，则

$$P(\text{用电 75 千瓦})=P(B_{10})=(0.2)^{10}=0.0000001,$$

$$P(\text{用电 67.5 千瓦})=P(B_9)=\binom{10}{9}\times(0.2)^9\times 0.8 =0.0000041,$$

$$P(\text{用电 60 千瓦})=P(B_8)=\binom{10}{8}\times(0.2)^8\times(0.8)^2 =0.0000737,$$

$$P(\text{用电 52.5 千瓦})=P(B_7)=\binom{10}{7}\times(0.2)^7\times(0.8)^3 =0.0007864,$$

$$P(\text{用电 45 千瓦})=P(B_6)=\binom{10}{6}\times(0.2)^6\times(0.8)^4 =0.005505.$$

所以，

$$P(\text{用电超过 48 千瓦})=P(B_{10})+P(B_9)+P(B_8)+P(B_7) =0.0008644\approx 1/1157.$$

结果表明，若供电 48 千瓦，超载的可能性只有 1/1157（即相当于 20 小时约有 1 分钟发生超载）。

在概率论中还讨论更一般 n 次独立试验，称为 n 次独立试验概型。n 重贝努里试验是 n 次独立试验概型的一种特殊情形。在 n 次独立试验概型中，不要求各次试验中成功的概率相同，每次试验也可以不只考虑两个可能结果。一般情况虽然复杂一些，但问题的解法还是类似的，第七节例 3、例 5 就是这种情形。

习 题 七

1. 10名射手命中率都是$\frac{1}{5}$，向同一目标彼此独立地各射击一次，试求：

(1) 10人中全没命中目标的概率；

(2) 恰有一人命中目标的概率；

(3) 至少有两人命中目标的概率.

2. 一工人看管三台机床，在一小时内机床不需要照顾的概率第一台是0.9，第二台是0.8，第三台是0.85，在一小时的过程中，试求：

(1) 没有一台机床需要照顾的概率；

(2) 至少有一台不需要照顾的概率.

3. 设在4次独立试验中，事件A出现的概率相等. 若已知事件A至少发生一次的概率等于65/81，求事件A在一次试验中出现的概率.

4. 设有两架高射炮，每一架击中飞机的概率都是0.6，试求同时射击一发炮弹而命中飞机的概率是多少？又若一架敌机侵犯，要以99%的概率击落它，问需要多少架高射炮？

5. 巴拿赫(Banach)问题：某数学家有两盒火柴，每一盒都装有n根. 每次使用时，他在任一盒中任取一根，问他发现一盒空而另一盒还有r根的概率多大？

*第九节 最简单流

一种观察对象在时间进程中随机地重复地出现，人们把这种现象称做流. 一个电话交换台收到的用户的呼唤形成一个呼唤流，到某商店去的顾客形成一顾客流，纺纱机上的断头形成断头流，放射性物质放出的质点形成质点流，等等. 流的例子不胜枚举. 在下面的叙述中，以呼唤流为代表，实际所指并不仅限于呼唤流. 在一个流中，对象的到来或疏或密带有

随机性，人们常常关心的是在一段时间内收到一定数量的呼唤(例如"收到 k 次呼唤")的概率.对于满足一定条件的流，这些事件的概率是可以确定出来的.

现在我们讨论具有下述性质的流.

(i) *平稳性* 在时间区间 $(a, a+t]$ 内收到 k 次呼唤的概率只与时间区间的长度 t 有关，而与时间的起点无关. 因此，我们可以用 $P_k(t)$ 表示在时间区间 $(a, a+t]$ 内收到 k 次呼唤的概率，而不需指明时间起点.

(ii) 在任一有限区间 $(a, a+t]$ 内只出现有限多次呼唤，

$$\sum_{k=0}^{\infty} P_k(t)=1,$$

且 $P_0(t)$ 不恒等于 1(否则意味着在任意长的时间区间内都实际上不出现呼唤，这种流显然没有什么实际意义).

(iii) *无后效性* 在不相交的时间区间内收到的呼唤次数相互独立，即对任意 n 个互不相交的区间 $(a_1, b_1]$, $(a_2, b_2]$, $\cdots$, $(a_n, b_n]$ 及任意非负整数 $k_1, k_2, \cdots, k_n$，事件 $B_{k_j}(a_j, b_j]$，"在 $(a_j, b_j]$ 内收到 k_j 次呼唤"$(j=1, 2, \cdots, n)$ 相互独立.

(iv) *普通性*

$$\sum_{j=2}^{\infty} P_j(\Delta t)=o(\Delta t),^{*)}$$

即

$$\lim_{\Delta t\to 0}\frac{\sum_{j=2}^{\infty} P_j(\Delta t)}{\Delta t}=0.$$

普通性的含义是当时间区间 $(t, t+\Delta t]$ 充分短时，发生 1 次以上呼唤的概率是 Δt 的高阶无穷小，因而在任一瞬间发生 1 次以上呼唤是实际不可能的.

---

*) 记号 $y=o(z)$ 表示 y 关于 z 是高阶无穷小，即 $y/z\to 0$. $y=g+o(z)$ 表示 $y-g=o(z)$. 可参考有关书籍，例如，复旦大学数学系编《数学分析》上册.

满足条件(i)～(iv)的流称为泊松(Poisson)流或最简单流.

下面我们对泊松流来确定出 $P_k(t)$.

首先, $P_0(t)$ 是 t 的单调下降函数, 并且由无后效性及平稳性

$$P_0(t+s)=P_0(t)P_0(s), \quad t, s\geqslant 0.$$

利用本节结尾给出的引理, 知

$$P_0(t)=\theta^t, \tag{1}$$

其中 $0\leqslant\theta\leqslant 1$.

注意 $\theta=P_0(1)\neq 1$. 否则, 由 (1), $P_0(t)\equiv 1$, 与条件(ii)矛盾. 同时还有 $\theta\neq 0$. 否则, 由(1), 对任意 $t>0$,

$$P_0(t)=0,$$

由此 $$P_1(t)=P_1\left(\frac{t}{2}\right)P_0\left(\frac{t}{2}\right)+P_0\left(\frac{t}{2}\right)P_1\left(\frac{t}{2}\right)=0,$$

$$P_2(t)=P_2\left(\frac{t}{2}\right)P_0\left(\frac{t}{2}\right)+P_1\left(\frac{t}{2}\right)P_1\left(\frac{t}{2}\right)$$

$$+P_0\left(\frac{t}{2}\right)P_2\left(\frac{t}{2}\right)=0,$$

....

因而 $$\sum_{k=0}^{\infty}P_k(t)=0, \quad t>0,$$

也与条件(ii)矛盾.

这样就知道 $0<\theta<1$, 故存在 $\lambda>0$, 使 $\theta=e^{-\lambda}$. 代入(1)得到

$$P_0(t)=e^{-\lambda t}. \tag{2}$$

现在来求 $P_k(t)$. 它是"在 $(0, t]$ 内出现 k 次呼唤"的概率. 将 $(0, t]$ 分成 n 个等长的互不相交的子区间, $n>k$, 设

A_1 表示事件"在某 k 个子区间中, 各恰好出现 1 次呼唤,

而在其余的 $n-k$ 个子区间内不出现呼唤”，

A_2 表示事件“至少有一子区间内出现一次以上呼唤”，

A_3 表示事件“在$(0, t]$内出现 k 次呼唤”，

则
$$P_k(t)=P(A_1)+P(A_2A_3). \tag{3}$$

分别讨论 $P(A_1)$ 和 $P(A_2A_3)$.

$$P(A_1)=\binom{n}{k}\left[P_1\left(\frac{t}{n}\right)\right]^k\left[P_0\left(\frac{t}{n}\right)\right]^{n-k},$$

而由(2)及普通性，当 $n\to\infty$ 时

$$\left[P_0\left(\frac{t}{n}\right)\right]^{n-k}=e^{-\lambda t\frac{n-k}{n}}=e^{-\lambda t}(1+o(1)),$$

$$\begin{aligned}\left[P_1\left(\frac{t}{n}\right)\right]^k&=\left[1-P_0\left(\frac{t}{n}\right)-\sum_{j=2}^{\infty}P_j\left(\frac{t}{n}\right)\right]^k\\&=\left[1-e^{-\lambda\frac{t}{n}}+o\left(\frac{t}{n}\right)\right]^k\\&=\left[\lambda\frac{t}{n}+o\left(\frac{t}{n}\right)\right]^k\\&=\frac{(\lambda t)^k}{n^k}[1+o(1)].\end{aligned}$$

故
$$P(A_1)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}[1+o(1)],$$

即
$$P(A_1)\to e^{-\lambda t}\frac{(\lambda t)^k}{k!}\quad(n\to\infty). \tag{4}$$

其次，$P(A_2A_3)\leqslant P(A_2)$，$P(A_2)$ 是 n 个子区间至少有一个出现 1 次以上呼唤的概率，而在一个指定的子区间内出现 1 次以上呼唤的概率为

$$1-P_0\left(\frac{t}{n}\right)-P_1\left(\frac{t}{n}\right)=\sum_{j=2}^{\infty}P_j\left(\frac{t}{n}\right)=o\left(\frac{t}{n}\right),$$

所以

$$P(A_2)\leqslant n\sum_{j=2}^{\infty}P_j\left(\frac{t}{n}\right)=t\frac{\sum\limits_{i=2}^{\infty}P_j\left(\frac{t}{n}\right)}{\frac{t}{n}}\to 0\quad(n\to\infty),$$

因而

$$P(A_2A_3)\to 0\quad(n\to\infty).\tag{5}$$

利用(3),(4),(5),最后我们得到

$$P_k(t)=\lim_{n\to\infty}(P(A_1)+P(A_2A_3))=e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

从所得的结果我们看到,在泊松流中,长为 t 的时间区间内发生 k 次呼唤的概率依赖于一个参数 λ,只要将 λ 的值确定出来,所求的概率就完全确定出来. λ 的意义将在以后看到(参看第三章,第一节,例 1),它表示在单位时间内收到呼唤的平均次数.

在确定 $P_k(t)$ 的过程中用到一个引理,现在叙述如下:

引理 若 $f(x)$ 是一连续函数(或单调函数),且对一切 x, y(或对一切 $x\geqslant 0$, $y\geqslant 0$)

$$f(x+y)=f(x)f(y)\tag{6}$$

则

$$f(x)=a^x,\tag{7}$$

其中 $a\geqslant 0$ 是某一常数.

【证】 由(6),

$$f(x)=\left[f\left(\frac{x}{2}\right)\right]^2\geqslant 0.$$

由(6)还容易导出

$$f(x_1+x_2+\cdots+x_n)=f(x_1)f(x_2)\cdots f(x_n).$$

于是记 $f(1)=a$,则有

$$f(1)=\left[f\left(\frac{1}{n}\right)\right]^n,\quad f\left(\frac{1}{n}\right)=a^{\frac{1}{n}}.$$

因而对任意正整数 m, n,

$$f\left(\frac{m}{n}\right)=\left[f\left(\frac{1}{n}\right)\right]^m=a^{\frac{m}{n}}.$$

这样就证明了(7)对一切正有理数成立.

对于任意实数 $x>0$, 取有理数序列 $\{r_n\}$, $\{s_n\}$,

$$r_1\leqslant r_2\leqslant\cdots\leqslant r_n\leqslant\cdots,$$

$$s_1\geqslant s_2\geqslant\cdots\geqslant s_n\geqslant\cdots,$$

$$0<r_n\leqslant x\leqslant s_n,\quad \lim_{n\to\infty}r_n=\lim_{n\to\infty}s_n=x.$$

当 $f(x)$ 是单调函数(不失一般性,设为单调上升函数)时,有

$$a^x=\lim_{n\to\infty}f(r_n)\leqslant f(x)\leqslant\lim_{n\to\infty}f(s_n)=a^x.$$

当 $f(x)$ 是连续函数时,有

$$f(x)=\lim_{n\to\infty}f(s_n)=a^x.$$

这样又证明了(7)对一切正实数成立.

由 (6), $f(0)=f(0)f(0)$, $f(0)=0$ 或 1. 前一情形下, $f(x)=f(x)f(0)=0$; 后一情况下,

$$f(0)=1=a^0,\ f(-x)=\frac{f(0)}{f(x)}=a^{-x}.$$

所以(7)获证. ▌

补 充 题

1. 把10本书任意放在书架上,求指定的3本书放在一起的概率.
2. 有7个球,7个盒子,每个球以相同的概率落入每个盒子,试求7个盒中球数依次为 2, 0, 3, 1, 0, 1, 0 的概率.
3. 一个分子从原点出发在数轴上游动四步,每次游动一个单位长,向左向右的机会相等,试求经过四步游动分子在原点的概率.
4. 袋中装有 l 个白球, m 个黑球及 n 个红球, 用返回方式取球 $l_1+m_1+n_1$ 次,试求

(1) 取出 l_1 个白球, m_1 个黑球, n_1 个红球的概率;

(2) 取出 l_1 个白球，m_1 个黑球，n_1 个红球，并且同一颜色的球连续出现的概率；

(3) 先取出 l_1 个白球，再取出 m_1 个黑球，最后取出 n_1 个红球的概率.

5. 从 6 双不同的手套中任取 4 只，试求恰有一双配对的概率.

6. 求 4 个随机数字中

(1) 全不相同的概率；

(2) 恰有两个相同的概率；

(3) 有三个相同的概率.

7. 袋中装有 m 个红球，n 个黑球 $(m\leqslant n)$，将球逐个取出，每次不返回，试求不连续出现两个红球的概率.

8. 袋中装有 m 个红球，n 个黑球，将球逐个取出，每次不返回，试求在第 r 次取到第 k 个红球的概率 $(1\leqslant k\leqslant m,\ 0\leqslant r-k\leqslant n)$.

9. 甲袋中装有 3 个红球，3 个白球，4 个黑球. 乙袋中装有 5 个红球，3 个白球，2 个黑球. 现从甲袋中任取 2 球，乙袋中任取 1 球，试求

(1) 3 个球颜色全同的概率；

(2) 3 个球颜色全不同的概率.

10. 在 N 个相同的球上分别写上号码 $1, 2, \cdots, N$，放在一个袋中. 现从袋中取 n 次球 $(n\leqslant N)$，每次任取 1 个，记录下号码然后放回. 试求依次记下的号码按严格上升次序排列的概率.

11. 在上题中，试求依次记下的号码按不降次序排列的概率.

12. 在第 10 题中对任何 n 及 $1\leqslant k\leqslant N$，求

(1) 记下的最大号码为 k 的概率；

(2) 记下的最小号码为 k 的概率.

13. 将 n 个球投入 N 个盒中 $(n\geqslant N)$，每个球都以 $\frac{1}{N}$ 的概率进每一个盒，试求没有空盒的概率.

14. 袋中有 n 个不同的球，用有返回方式抽球一直到每个球至少被抽一次. 试求抽球次数为 $N(N\geqslant n)$ 的概率.

15. 甲、乙两人轮流射击，先射中目标的为胜. 已知甲射中目标的概率为 α，乙射中目标的概率为 β. 甲先射，试求甲胜的概率，乙胜的

概率.

16. 甲和乙比赛射击,每进行一次胜者得一分. 在一次射击中, 甲胜的概率是 α, 乙胜的概率是 β. 射击进行到有一人超过对方 2 分就停止,多得 2 分者为胜. 试分别求甲获胜的概率和乙获胜的概率.

17. 甲、乙、丙三人按下面规则进行比赛,第一局由甲、乙参加, 丙轮空, 第二局由第一局的优胜者和丙参加,第一局的失败者轮空, 比赛用这种方式一直进行到其中一个人连胜两局为止, 连胜两局者为整场比赛的优胜者. 若甲、乙、丙胜每局的概率各为 $\frac{1}{2}$, 问甲、乙、丙成为整场优胜者的概率各是多少?

18. 从数轴上随机地取一点, ξ 表示这个点的坐标. 对任何实数 x, 用 A_x 表示事件 $\{\xi<x\}$, 试用事件 $A_x(-\infty<x<+\infty)$ 表示下列事件:

$\{a\leqslant\xi<b\}$, $\{\xi\leqslant b\}$, $\{a<\xi<b\}$, $\{\xi<+\infty\}$, $\{\xi\geqslant a\}$,

其中 a, b 是任意指定的实数.

第一章小结

概率论的研究对象是随机试验中的随机现象. 概率论的基本概念有着丰富的现实背景, 所以结合直观含义来理解概念非常重要.

重复试验是认识随机现象规律的重要途径之一. 在重复试验中事件出现的频率是一个重要概念, 它为概率概念的形成提供了一个直观依据.

样本空间 Ω, 事件 (事件域 $\mathscr{F}$) 与概率 P 是概率论中最基本的概念. 在概率论中用 $(\Omega, \mathscr{F}, P)$——称为概率空间, 来描述随机现象. 概率论的一般理论就是在一个任意给定的概率空间上展开的. 分析具体问题时需要明确样本空间, 事件域, 并确定出概率. 但在许多场合下我们假定有关这个具

体问题的概率空间存在，这时，也可以不指明概率空间的具体结构而讨论某些性质或计算某些概率．如第六节中推导全概率公式，第九节中确定 $P_k(t)$ 的形式等等．

概率的公理化定义规定了概率的基本性质．不同的随机试验中，概率的具体形式和数值不同，需结合随机现象的特点来确定．在本章中讨论了四种常见的类型：古典型、几何型、n 次独立试验及泊松流，确定出这些随机现象中事件的概率．这四种类型的随机试验概括了相当广泛的实际问题，具有重要的应用．

独立性概念在概率论中具有重要地位．n 次独立试验概型就是基于独立性概念建立起来的．对于独立试验的研究，在概率论中有许多重要的结论，在第四章中可以看到一些．

引入条件概率可以使我们更好地理解独立性的含义．另外，当讨论不具有独立性的现象时，条件概率是一个重要工具，例如用概率的乘法公式可以解决许多概率计算的问题．

这一章是全书的基础．事件，概率，事件的独立性这些基本概念将贯穿全书各章节，学过这一章后应当熟悉事件的运算，概率的性质，应当理解事件独立的含义．

第二章

随机变量

第一节　随机变量及其分布函数

在随机试验中，我们的观察对象常常是一个或若干个随机取值的量，观察的目的是要了解它们取各可能值或在各个范围内取值的概率．在第一章中我们已经见到许多这样的问题．在产品的抽样检查中，我们要观察的是抽出的样品中次品的个数 ξ，并且希望知道，"ξ 取值 k"，"ξ 取小于 k 的值"($k=0, 1, 2, \cdots$)等事件的概率．在 n 次独立试验中成功的次数 μ 也是一个随机取值的量，我们所关心的是"μ 取值 k"以及"μ 取小于 k 的值"，"μ 取大于 k 的值"($k=0, 1, 2, \cdots$)等事件的概率．在几何概型中，投点的坐标(ξ_1, ξ_2)也是随机取值的量，在这种试验中，我们希望知道"(ξ_1, ξ_2)落在区域 A 内"的概率．由此看来，研究随机试验中随机取值的量具有重要意义．

所谓随机试验中随机取值的量 ξ，确切地说，指的是样本空间 Ω 上的一个函数 $\xi=\xi(\omega)$，即对应于每一样本点 $\omega\in\Omega$，ξ 取一确定的实数值 $\xi(\omega)$．在一次试验中，若出现了样本点 ω，则 ξ 就取值 $\xi(\omega)$；由于在一次试验中究竟出现哪一个样本点带有随机性，因而 ξ 的取值也就带有随机性．对于样本空间上的函数 ξ，我们关心的是它在一定的范围内取值的事件，这些事件可以明确地写出它们的集合表示：

$$\{\xi=a\}=\{\omega|\xi(\omega)=a\},$$

$$\{\xi<x\}=\{\omega|\xi(\omega)<x\},$$
$$\{a\leqslant\xi<b\}=\{\omega|a\leqslant\xi(\omega)<b\}$$

等等,这些事件的概率记作

$$P(\xi=a),\quad P(\xi<x),\quad P(a\leqslant\xi<b).$$

这些概率描述了 ξ 取值的规律,因而掌握所有这些概率就成为我们研究的中心任务.

为了熟悉上面引出的概念,我们来看几个简单的例子.

【例 1】 有五件产品,其中两件是次品(用 a_1, a_2 表示),三件是正品(用 b_1, b_2, b_3 表示),从中任意取出两件,这个随机试验的样本空间为

$$\begin{aligned}\Omega=\{&\{a_1, a_2\}, \{a_1, b_1\}, \{a_1, b_2\}, \{a_1, b_3\},\\ &\{a_2, b_1\}, \{a_2, b_2\}, \{a_2, b_3\},\\ &\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}.\end{aligned}$$

考虑抽取的两件产品中次品的件数 ξ,则 ξ 是 Ω 上的一个函数 $\xi=\xi(\omega)$,这个函数可具体写出如下:

$$\xi(\omega)=\begin{cases}0, & \text{当 } \omega=\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\};\\ 1, & \text{当 } \omega=\{a_1, b_1\}, \{a_1, b_2\}, \{a_1, b_3\},\\ & \quad\{a_2, b_1\}, \{a_2, b_2\}, \{a_2, b_3\};\\ 2, & \text{当 } \omega=\{a_1, a_2\},\end{cases}$$

在一次试验中:若出现了样本点 $\{a_1, b_3\}$,则 ξ 就取值 1;若出现了样本点 $\{b_2, b_3\}$,则 ξ 就取值 0;一般,若出现了样本点 ω,则 ξ 就取对应的值 $\xi(\omega)$. 由于在每次试验中究竟出现哪一个样本点带有随机性,所以,ξ 的取值也就带有随机性.

ξ 取各个可能值,以及 ξ 在一定的范围内取值等事件用符号表示如下:

$$\begin{aligned}\{\xi=0\}&=\{\omega|\xi(\omega)=0\}\\ &=\{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\},\end{aligned}$$

$$\{\xi=1\}=\{\omega|\xi(\omega)=1\}$$
$$=\{\{a_1, b_1\}, \{a_1, b_2\}, \{a_1, b_3\},$$
$$\{a_2, b_1\}, \{a_2, b_2\}, \{a_2, b_3\}\},$$
$$\{\xi=2\}=\{\omega|\xi(\omega)=2\}=\{\{a_1, a_2\}\}.$$

容易求出它们的概率

$$P(\xi=0)=\frac{3}{10}, \quad P(\xi=1)=\frac{6}{10}, \quad P(\xi=2)=\frac{1}{10}.$$

在应用中,我们还需要知道 $\{\xi<0\}$, $\{\xi<1\}$, $\{\xi<2\}$, $\{\xi<3\}$ 等事件的概率,注意到

$$\{\xi<0\}=\emptyset, \quad \{\xi<1\}=\{\xi=0\},$$
$$\{\xi<2\}=\{\xi=0\}\cup\{\xi=1\},$$
$$\{\xi<3\}=\{\xi=0\}\cup\{\xi=1\}\cup\{\xi=2\},$$

所以,

$$P(\xi<0)=0, \quad P(\xi<1)=\frac{3}{10},$$
$$P(\xi<2)=\frac{9}{10}, \quad P(\xi<3)=1.$$

【例 2】 在 n 次独立试验中,令 ξ 为 n 次试验中成功的次数,则 ξ 也是样本空间上的函数. 以 $n=4$ 的情形为例,样本空间

$$\begin{aligned}\Omega=\{&(0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 0),\\ &(0, 1, 0, 0), (1, 0, 0, 0), (0, 0, 1, 1),\\ &(0, 1, 0, 1), (0, 1, 1, 0), (1, 0, 0, 1),\\ &(1, 0, 1, 0), (1, 1, 0, 0), (0, 1, 1, 1),\\ &(1, 0, 1, 1), (1, 1, 0, 1), (1, 1, 1, 0),\\ &(1, 1, 1, 1)\}\end{aligned}$$

(参看第一章,第八节,例 1). 函数 $\xi=\xi(\omega)$ 为

$$\xi(\omega)=\begin{cases}0, & \text{当 } \omega=(0,0,0,0);\\ 1, & \text{当 } \omega=(0,0,0,1),(0,0,1,0),\\ & \qquad (0,1,0,0),(1,0,0,0);\\ 2, & \text{当 } \omega=(0,0,1,1),(0,1,0,1),\\ & \qquad (0,1,1,0),(1,0,0,1),\\ & \qquad (1,0,1,0),(1,1,0,0);\\ 3, & \text{当 } \omega=(0,1,1,1),(1,0,1,1),\\ & \qquad (1,1,0,1),(1,1,1,0);\\ 4, & \text{当 } \omega=(1,1,1,1).\end{cases}$$

我们感兴趣的事件是:

$$\{\xi=0\}=\{\omega|\xi(\omega)=0\}=\{(0,0,0,0)\},$$

$$\{\xi=1\}=\{\omega|\xi(\omega)=1\}$$
$$=\{(0,0,0,1),(0,0,1,0),(0,1,0,0),$$
$$(1,0,0,0)\},$$

$$\{\xi=2\}=\{\omega|\xi(\omega)=2\}$$
$$=\{(0,0,1,1),(0,1,0,1),(0,1,1,0),$$
$$(1,0,0,1),(1,0,1,0),(1,1,0,0)\},$$

$$\{\xi=3\}=\{\omega|\xi(\omega)=3\}$$
$$=\{(0,1,1,1),(1,0,1,1),(1,1,0,1),$$
$$(1,1,1,0)\},$$

$$\{\xi=4\}=\{\omega|\xi(\omega)=4\}=\{(1,1,1,1)\}.$$

如果在每次试验中成功的概率为 p, 则按 n 重贝努里试验知上述事件的概率为

$$P(\xi=k)=\binom{4}{k}p^k(1-p)^{4-k}, \quad k=0,\ 1,\ 2,\ 3,\ 4.$$

在实际应用中，常常还需要知道 ξ 在某一范围内取值的概率，如

$$P(\xi<0)=P(\emptyset)=0,$$

$$P(\xi<1)=P(\xi=0)=(1-p)^4,$$

$$P(\xi<2)=\sum_{k=0}^{1}P(\xi=k)=\sum_{k=0}^{1}\binom{4}{k}p^k(1-p)^{4-k},$$

$$P(\xi<3)=\sum_{k=0}^{2}(\xi=k)=\sum_{k=0}^{2}\binom{4}{k}p^k(1-p)^{4-k},$$

$$P(\xi<4)=\sum_{k=0}^{3}(\xi=k)=\sum_{k=0}^{3}\binom{4}{k}p^k(1-p)^{4-k},$$

$$P(\xi<5)=\sum_{k=0}^{4}(\xi=k)=\sum_{k=0}^{4}\binom{4}{k}p^k(1-p)^{4-k}=1.$$

此外，还可以利用事件的关系和概率性质得出 ξ 落在其它形式的范围内的概率．如由

$$\{\xi\geqslant 2\}=\overline{\{\xi<2\}}\quad（对立事件），$$

可得

$$P(\xi\geqslant 2)=1-P(\xi<2),$$

又由

$$\{1\leqslant\xi<3\}=\{\xi<3\}-\{\xi<1\},\ \{\xi<1\}\subseteq\{\xi<3\},$$

可得

$$P(1\leqslant\xi<3)=P(\xi<3)-P(\xi<1).$$

【例 3】 在 n 重贝努里试验中，也常常考虑其它的随机取值的量，如第 1 次试验中成功的次数 μ_1，这也是样本空间上的一个函数，仍以 $n=4$ 的情形为例，这个函数具体写出就是

$$\mu_1(\omega)=\begin{cases}0, & \text{当 } \omega=(0,0,0,0),(0,0,0,1),\\ & (0,0,1,0),(0,1,0,0),\\ & (0,0,1,1),(0,1,0,1),\\ & (0,1,1,0),(0,1,1,1);\\ 1, & \text{当 } \omega=(1,0,0,0),(1,0,0,1),\\ & (1,0,1,0),(1,1,0,0),\\ & (1,0,1,1),(1,1,0,1),\\ & (1,1,1,0),(1,1,1,1).\end{cases}$$

事件$\{\mu_1=0\}$表示"第1次试验中没成功"，事件$\{\mu_1=1\}$表示"第1次试验中成功"，它们的概率为

$$P(\mu_1=1)=p,\quad P(\mu_1=0)=1-p.$$

【例 4】 在会面问题中，样本空间为

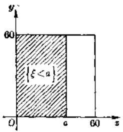


图 2-1

$$\Omega=\{(x,\ y)\,|\,0\leqslant x\leqslant 60,\ 0\leqslant y\leqslant 60\}.$$

设ξ，η分别表示甲、乙两人到达的时刻，则它们都是样本空间上的函数：

$$\xi(x,\ y)=x,\quad \eta(x,\ y)=y,$$
$$(x,\ y)\in\Omega.$$

事件

$$\{\xi<a\}=\{(x,\ y)\,|\,\xi(x,\ y)<a\}$$
$$=\{(x,\ y)\,|\,0\leqslant x<a,\ 0\leqslant y\leqslant 60\},$$

其中$0<a\leqslant 60$. 按几何概型，

$$P(\xi<a)=\frac{\text{区域}\{(x,\ y)\,|\,0\leqslant x<a,\ 0\leqslant y\leqslant 60\}\text{的面积}}{\Omega\text{ 的面积}}$$
$$=\frac{a}{60}.$$

当$a\leqslant 0$时，$\{\xi<a\}=\emptyset$，$P(\xi<a)=0$.

当 $a>60$ 时，$\{\xi<a\}=\Omega$，$P(\xi<a)=1$.

类似地，

$$P(\eta<b)=\begin{cases}0, & \text{当 } b\leqslant 0;\\ \dfrac{b}{60}, & \text{当 } 0<b\leqslant 60;\\ 1, & \text{当 } b>60.\end{cases}$$

在会面问题中，不仅需要知道 ξ，η 单独在一定区间内取值的概率，还要知道它们合在一起在平面区域内取值的概率，我们将在第四节中讨论.

定义 设$(\Omega, \mathscr{F}, P)$是一给定的概率空间，若 $\xi=\xi(\omega)$ 是样本空间上的一个实值函数，且对任一实数 x，

$$\{\xi<x\}=\{\omega|\xi(\omega)<x\}\in\mathscr{F}$$

(即$\{\xi<x\}$是事件)，则称 ξ 是$(\Omega, \mathscr{F}, P)$上的一个**随机变量**. 对任意实数 x，事件$\{\xi<x\}$的概率

$$F(x)=P(\xi<x)$$

称为随机变量 ξ 的**分布函数**.

在上面的例 1～4 中，我们已经举出不少随机变量的实例. 同时，也可以看到第一章中讨论的问题都可以归结为对随机变量的研究. 事实上，当我们要讨论概率空间$(\Omega, \mathscr{F}, P)$上的某一事件 A 时，我们可以引进一个随机变量 $\xi_A=\xi_A(\omega)$：

$$\xi_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A.\end{cases}$$

于是事件 A 的概率便等于 ξ_A 取值 1 的概率

$$P(A)=P(\xi_A=1).$$

这样，对任意事件 A 的研究都可以归结为对随机变量的研究. 因此，我们将随机变量作为今后研究的主要对象.

为了理解分布函数的概念，我们将前面例 1～4 中的随机变量的分布函数分别求出如下.

在例 1 中，

当 $x\leqslant 0$ 时，$P(\xi<x)=P(\emptyset)=0$；

当 $0<x\leqslant 1$ 时，$P(\xi<x)=P(\xi=0)=\frac{3}{10}$；

当 $1<x\leqslant 2$ 时，$P(\xi<x)=P(\xi=0)+P(\xi=1)=\frac{9}{10}$；

当 $x>2$ 时，$P(\xi<x)=P(\Omega)=1$.

所以，我们得到 ξ 的分布函数为

$$F(x)=\begin{cases}0, & \text{当 } x\leqslant 0;\\ \frac{3}{10}, & \text{当 } 0<x\leqslant 1;\\ \frac{9}{10}, & \text{当 } 1<x\leqslant 2;\\ 1, & \text{当 } x>2.\end{cases}$$

类似地，可求出例 2 中随机变量 ξ 的分布函数为

$$F(x)=\begin{cases}0, & \text{当 } x\leqslant 0;\\ (1-p)^4, & \text{当 } 0<x\leqslant 1;\\ \sum\limits_{k=0}^{1}\binom{4}{k}p^k(1-p)^{4-k}, & \text{当 } 1<x\leqslant 2;\\ \sum\limits_{k=0}^{2}\binom{4}{k}p^k(1-p)^{4-k}, & \text{当 } 2<x\leqslant 3;\\ \sum\limits_{k=0}^{3}\binom{4}{k}p^k(1-p)^{4-k}, & \text{当 } 3<x\leqslant 4;\\ 1, & \text{当 } x>4.\end{cases}$$

在例 3 中，随机变量 μ_1 的分布函数为

$$F(x)=\begin{cases}0, & 当\ x\leqslant 0;\\ 1-p, & 当\ 0<x\leqslant 1;\\ 1, & 当\ x>1.\end{cases}$$

在例 4 中，随机变量 ξ 的分布函数为

$$F(x)=\begin{cases}0, & 当\ x\leqslant 0;\\ \dfrac{x}{60}, & 当\ 0<x\leqslant 60;\\ 1, & 当\ x>60.\end{cases}$$

研究随机变量 ξ，就是要掌握 ξ 在各个范围内取值的概率．上面定义中引入的分布函数 $F(x)$ 就是掌握 ξ 取值规律的一个基本工具．

由定义，知道了分布函数，就知道了 ξ 落在区间 $(-\infty, x)$ 内的概率．由此，又可知 ξ 落在任意有限区间 $[a, b)$ $(a<b)$ 的概率 $P(a\leqslant\xi<b)$．因为

$$\{a\leqslant\xi<b\}=\{\xi<b\}-\{\xi<a\},\ \{\xi<a\}\subseteq\{\xi<b\}$$

故有

$$P(a\leqslant\xi<b)=P(\xi<b)-P(\xi<a)=F(b)-F(a). \tag{1}$$

下面我们还将说明利用分布函数可以得出 ξ 落在其它形式的范围内的概率．在此我们先来给出分布函数的基本性质．

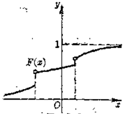


图 2-2

1) $F(x)$ 是 x 的不降函数：若 $a<b$，则 $F(a)\leqslant F(b)$．

2) $0\leqslant F(x)\leqslant 1$，　$F(-\infty)=0$，$F(+\infty)=1$，

其中 $F(-\infty)=\lim\limits_{x\to-\infty}F(x)$，　$F(+\infty)=\lim\limits_{x\to+\infty}F(x)$．

3) $F(x)$ 在任一点左连续：$F(x-0)=F(x)$，

其中 $F(x-0)=\lim\limits_{y\to x-0}F(y)$.

基本性质的证明如下.

1) 可由(1)得出(利用概率的非负性).

2) 中第一式是显然的. 由于 $F(x)$ 单调有界, $\lim\limits_{x\to-\infty}F(x)$ 及 $\lim\limits_{x\to+\infty}F(x)$ 都存在, 并且

$$\lim_{x\to+\infty}F(x)=\lim_{n\to+\infty}F(n),$$
$$\lim_{x\to-\infty}F(x)=\lim_{n\to+\infty}F(-n).$$

考虑事件关系

$$\{\xi<1\}\subseteq\{\xi<2\}\subseteq\cdots\subseteq\{\xi<n\}\subseteq\cdots,$$
$$\bigcup_{n=1}^{+\infty}\{\xi<n\}=\Omega,$$

利用概率的下连续性(第一章第五节性质 9)得

$$\lim_{n\to+\infty}F(n)=\lim_{n\to+\infty}P\{\xi<n\}=P\left(\bigcup_{n=1}^{\infty}\{\xi<n\}\right)=P(\Omega)=1.$$

2)中第三式获证. 又

$$\{\xi<-1\}\supseteq\{\xi<-2\}\supseteq\cdots\supseteq\{\xi<-n\}\supseteq\cdots,$$
$$\bigcap_{n=1}^{+\infty}\{\xi<-n\}=\emptyset,$$

利用概率的上连续性(第一章第五节性质 8)得

$$\lim_{n\to+\infty}F(-n)=\lim_{n\to+\infty}P(\xi<-n)=P\left(\bigcap_{n=1}^{+\infty}\{\xi<-n\}\right)$$
$$=P(\emptyset)=0.$$

这样又证明了 2) 中第二式.

3)的证明. 仍根据 $F(x)$ 的单调有界性质, 对任何 x, $\lim\limits_{y\to x-0}F(y)$ 存在, 并且

$$\lim_{y\to x-0}F(y)=\lim_{n\to+\infty}F\left(x-\frac{1}{n}\right).$$

考虑事件关系

$$\{\xi<x-1\}\subseteq\left\{\xi<x-\frac{1}{2}\right\}\subseteq\cdots\subseteq\left\{\xi<x-\frac{1}{n}\right\}\subseteq\cdots,$$

$$\bigcup_{n=1}^{+\infty}\left\{\xi<x-\frac{1}{n}\right\}=\{\xi<x\},$$

利用概率的下连续性(第一章第五节性质9)得

$$\lim_{n\to+\infty}F\left(x-\frac{1}{n}\right)=\lim_{n\to+\infty}P\left(\xi<x-\frac{1}{n}\right)$$

$$=P\left(\bigcup_{n=1}^{+\infty}\left\{\xi<x-\frac{1}{n}\right\}\right)=P(\xi<x)=F(x).$$

现在利用分布函数再给出几个公式，用来计算 ξ 在某些范围内取值的概率.

利用对立事件 $\{\xi\geqslant x\}=\overline{\{\xi<x\}}$，可得

$$P(\xi\geqslant x)=1-F(x). \tag{2}$$

但计算 $\{\xi\leqslant x\}$ 的概率却需要利用下述公式

$$P(\xi\leqslant x)=F(x+0),\quad F(x+0)=\lim_{y\to x+0}F(y). \tag{3}$$

这个公式可以这样推导

$$P(\xi\leqslant x)=P\left(\bigcap_{n=1}^{+\infty}\left\{\xi<x+\frac{1}{n}\right\}\right)=\lim_{n\to+\infty}P\left(\xi<x+\frac{1}{n}\right)$$

$$=\lim_{n\to+\infty}F\left(x+\frac{1}{n}\right)=\lim_{y\to x+0}F(y).$$

由(3)立即又得

$$P(\xi>x)=1-P(\xi\leqslant x)=1-F(x+0), \tag{4}$$

$$P(\xi=x)=P(\xi\leqslant x)-P(\xi<x)=F(x+0)-F(x), \tag{5}$$

$$P(a\leqslant\xi\leqslant b)=F(b+0)-F(a), \tag{6}$$

$$P(a<\xi<b)=F(b)-F(a+0), \tag{7}$$

$$P(a<\xi\leqslant b)=F(b+0)-F(a+0). \tag{8}$$

通过以上的讨论，我们看到作为一个分布函数必需具有

基本性质 1)～3)，同时也了解到掌握了分布函数可以进一步得出 ξ 落在各种形式区域内的概率。更一般地，有下述重要结论：

一、若 $F(x)$ 具有基本性质 1)～3)（即单调不降；$F(-\infty)=0$，$F(+\infty)=1$；且左连续），则存在某一概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变量 ξ，它的分布函数恰好就是 $F(x)$。

因此，作为一个分布函数，必需也只需满足 1)～3)。

二、对于任一波雷尔集 B 来说，$P(\xi\in B)$ 由 ξ 的分布函数 $F(x)$ 唯一地决定。

因此，知道了分布函数，也就相当于知道了全部 $P(\xi\in B)$。在概率论中，称 $P(\xi\in B)$（对一切波雷尔集 B）为 ξ 的概率分布（或分布律），并用来作为 ξ 的取值规律的完全的描述。如上所述，分布函数也可作为对 ξ 取值规律的完全描述。所以，今后我们将通过分布函数来掌握 ξ 的取值规律。

这些问题的确切叙述和详细的讨论超出了本书的范围，有兴趣的读者可参考有关书籍。

习 题 一

1. 投掷两次骰子*)，试求出现点数之和 ξ 的分布函数。
2. 从 1, 2, 3, 4, 5 中任意取出两个数码，设 ξ 是这两个数码中的最大数码，试求 ξ 的分布函数。
3. 将红、绿、白三个球任意放到编号为 1, 2, 3 的三个盒中，设 ξ 是放入 1 盒中球的个数，η 是放有红球的盒子号码，试分别求出 ξ 和 η 的分布函数。
4. 设在每次试验中，ξ 永远取常数 c，问 ξ 是不是随机变量，如果是，它的分布函数是什么？
5. 设 ξ 是一随机变量，已知它的分布函数为 $F(x)$，试用分布函数给

*) 骰子是一个骨制的小正六面体，各面分别刻有 1～6 的点数。

出下列事件的概率:

$$\{|\xi|<1\},\ \{|\xi-2|<3\},\ \{2\xi+1>5\},$$
$$\{\xi^2\leqslant 4\},\ \{\xi^3<8\}.$$

6. 设 ξ 是一随机变量,定义

$$G(x)=P(\xi\leqslant x),$$

试证: (i) $G(x)$ 非降, (ii) $G(-\infty)=0$, $G(+\infty)=1$, (iii) $G(x)$ 在每一点右连续.

第二节 离散型随机变量

2.1 离散型随机变量及其分布列

只取有限个或可数个不同可能值的随机变量, 称为离散型随机变量.

从一批产品中抽取 n 件, 抽到的次品个数 ξ 只能取有限个可能值 0, 1, ⋯, n; n 次独立试验中成功的次数 μ 只能取有限个可能值 0, 1, ⋯, μ; 某交换台在单位时间内收到的呼唤次数 ξ 只能取可数个可能值 0, 1, 2, ⋯, k, ⋯. 以上所说的这些随机变量都是离散型的随机变量.

设 ξ 是一离散型随机变量, 它所取的全部可能值为 x_1, x_2, ⋯, x_k, ⋯, 则 ξ 取各可能值的概率

$$p_k=P(\xi=x_k),\quad k=1,\ 2,\ \cdots,$$

称为随机变量 ξ 的分布列. 离散型随机变量的分布列常用下述形式表示:

ξ	x_1	x_2	⋯	x_k	⋯
P	p_1	p_2	⋯	p_k	⋯

或

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}.$$

离散型随机变量的分布列具有下列性质：

1) $p_k \geqslant 0, \quad k=1, 2, \cdots$；

2) $\sum\limits_{k=1}^{\infty} p_k = 1$.

这两个性质容易利用概率的性质证明. 因为概率值是非负实数，所以 $p_k = P(\xi = x_k) \geqslant 0$，即 1)成立. 由于 $x_1, x_2, \cdots, x_k, \cdots$是 ξ 所取的全部可能值，因此

$$\Omega = \{\xi = x_1\} \cup \{\xi = x_2\} \cup \cdots \cup \{\xi = x_k\} \cup \cdots.$$

注意到右端各事件两两互不相容，利用概率的可加性即得

$$1 = P(\xi = x_1) + P(\xi = x_2) + \cdots + P(\xi = x_k) + \cdots,$$

这样又证明了 2).

对于离散型随机变量 ξ 来说，由于它所取的可能值至多有可数个，所以，它在某一范围内取值的概率等于它取这个范围内的各可能值的概率之和，特别，离散型随机变量 ξ 的分布函数可以利用它的分布列得出

$$P(\xi < x) = \sum_{x_k < x} P(\xi = x_k),$$

即

$$F(x) = \sum_{x_k < x} p_k.$$

因而 ξ 的分布列可以唯一决定其概率分布.

前节例 1 中，随机变量 ξ 的分布列为

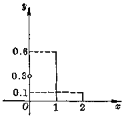


图 2-3

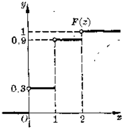


图 2-4

$$\begin{pmatrix} 0 & 1 & 2 \\ 0.3 & 0.6 & 0.1 \end{pmatrix},$$

分布函数为 $$F(x)=\begin{cases} 0, & x\leqslant 0; \\ 0.3, & 0<x\leqslant 1; \\ 0.9, & 1<x\leqslant 2; \\ 1, & x>2. \end{cases}$$

离散型随机变量的分布列和分布函数可利用图形来示意，参看图2-3，2-4．离散型随机变量的分布函数是一阶梯形函数．

知道了离散型随机变量的分布列，便可以知道它在任意范围内取值的概率，它和分布函数具有同样的作用；但分布列更为直接而简便，所以，对于离散型随机变量，将通过它的分布列来掌握它的概率规律．

2.2 几个常见的离散型分布

不同的随机变量具有不同的概率分布，现在，我们讨论几个具有广泛应用的离散型概率分布．

1．超几何分布

一批产品共 N 件，其中有 M 件次品，从中任意取出 n 件取到的次品件数 ξ 是一离散型随机变量，其概率分布为(参看第一章，第三节，例1)

$$P(\xi=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k=0,\ 1,\ \cdots,\ n,^{*)} \quad (1)$$

*) 若 $r<s$，规定 $\binom{r}{s}=0$．

这个概率分布通常称为超几何分布．当一个随机变量 ξ 的概率分布为超几何分布时，我们常常说 ξ 服从超几何分布，或说 ξ 具有超几何分布．

2．二项分布

在 n 重贝努里试验中，设每次试验某事件 A 出现(即"成功")的概率为 p，令 ξ 是 n 次试验中事件 A 出现的次数，则 ξ 是一个离散型随机变量，它的概率分布(参看第一章第八节)为

$$P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,\ 1,\ \cdots,\ n. \tag{2}$$

这个概率分布通常称为二项分布(因由二项式 $(p+q)^n$ 展开的各项组成)．二项分布中含有两个参数 n，p，当它们的值已知时，便可计算出分布列中的各概率值．参数为 n，p 的二项分布简记作 $b(n,\ p)$．当一个随机变量 ξ 的概率分布为二项分布时，称 ξ 服从二项分布，或 ξ 具有二项分布，用符号记作

$$\xi\sim b(n,\ p).$$

一种概率分布所以重要，是因为它具有广泛的现实背景，因而具有广泛的应用．在第一章中我们已经了解到 n 重贝努里试验的广泛背景，因此二项分布的重要性也就十分明显了．

一种概率分布所以重要，还因为它常常是另一些分布的极限．现在我们来说明二项分布是超几何分布的极限，确切地说，有下面的极限定理．

定理 设在超几何分布中，n 是一个取定的正整数，而

$$\lim_{N\to\infty}\frac{M}{N}=p,$$

则

$$\lim_{N\to\infty}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,\ 1,\ \cdots,\ n.$$

(3)

【证】 $\binom{M}{k}\binom{N-M}{n-k}\Big/\binom{N}{n}$

$$=\frac{n!}{k!(n-k)!}M(M-1)\cdots(M-k+1)$$
$$\times(N-M)(N-M-1)\cdots(N-M-n+k+1)$$
$$\times[N(N-1)\cdots(N-n+1)]^{-1}$$
$$=\binom{n}{k}\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\left(\frac{M}{N}-\frac{k-1}{N}\right)$$
$$\times\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\cdots\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)$$
$$\times\left[1\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\right]^{-1}$$
$$\to\binom{n}{k}p^k(1-p)^{n-k},\quad N\to\infty.$$ ▌

(3)式表明，当 N 充分大时，超几何分布可用二项分布来近似代替．这个结论从直观上来看也是很清楚的．二项分布可以用来描述返回抽样，而超几何分布可以用来描述不返回抽样，虽然二者并不相同，但当 N(抽样对象的总数)很大时，两种抽样的差别是不会很大的．

3．泊松分布

如下形式的分布列

$$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,\ 1,\ \cdots,\ n,\ \cdots,\qquad(4)$$

(其中 λ 是一正实数，e 是自然对数的底）称为泊松分布.

从第一章第九节我们知道，若 ξ 是泊松流在某一时间区间内出现的质点个数，则 ξ 具有泊松分布，可见泊松分布是具有广泛实际背景和实际应用的一种概率分布.

同时，泊松分布还是二项分布的极限分布，这可以从下面的极限定理看出.

定理 设 λ 是一正常数，$p_n=\frac{\lambda}{n}$，则有

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,\ 1,\ 2,\ \cdots. \qquad (5)$$

【证】

$$\begin{aligned}\binom{n}{k}p_n^k(1-p_n)^{n-k}&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\frac{\lambda^k}{k!}\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}.\end{aligned}$$

对任意固定的非负整数 k，

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda},$$

这样我们就证明了(5)式. ▌

利用上面的极限定理，可以看出，对二项分布 $b(n,\ p)$ 来说，若 n 充分大，p 相对很小，则

$$P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,\ 1,\ \cdots,\ n,$$

其中 $\lambda=np$. 这就是说，当 n 充分大时，二项分布 $b(n, p)$ 可近似地看作是以 $\lambda=np$ 为参数的泊松分布. 为了直观地了解二者的关系，可以参看下面的数值计算结果的对照表：

表 2-1

k	二项分布 $\binom{n}{k}p^k(1-p)^{n-k}$				泊松分布 $\frac{\lambda^k}{k!}e^{-\lambda}$
	$n=10$ $p=0.1$	$n=20$ $p=0.05$	$n=40$ $p=0.025$	$n=100$ $p=0.01$	$\lambda=1\ (=np)$
0	0.349	0.358	0.363	0.366	0.368
1	0.385	0.377	0.372	0.370	0.368
2	0.194	0.189	0.186	0.185	0.184
3	0.057	0.060	0.060	0.061	0.061
4	0.011	0.013	0.014	0.015	0.015
⋮	⋮	⋮	⋮	⋮	⋮

4．几何分布

设某射手在一次射击中的命中率为 p，今向一目标连续射击直到击中目标为止，试求到击中目标时所需的射击次数 ξ 的分布列.

ξ 所取的可能值为 $1, 2, \cdots, k, \cdots$.

$\{\xi=k\}$ 相当于"第一次射击未中，…，第 $k-1$ 次射击未中，第 k 次射击命中"，由于各次射击是独立进行的，而每次射击命中的概率为 p，不中的概率为 $q=1-p$，故可得

$$P(\xi=k)=pq^{k-1}, \quad k=1, 2, \cdots, \tag{6}$$

这样我们就确定出 ξ 的分布列.（容易验证(6)中的序列具有分布列的基本性质 1)，2).）

(6)中的分布列称为几何分布，它所描述的是独立重复试验中首次成功的试验次数的概率分布，在实际应用中经常遇到.

习 题 二

1. 设 ξ 的分布列中含有一未知常数 C，试将 C 确定出来：

   (1) $P(\xi=k)=\frac{C}{N}$, $\quad k=1, 2, \cdots, N$;

   (2) $P(\xi=k)=C\frac{\lambda^k}{k!}$, $\quad k=1, 2, \cdots$;

   (3) $P(\xi=k)=C\frac{1}{2^k}$, $\quad k=1, 2, \cdots$.

2. 试写出几何分布的分布函数.
3. 设有某一类似于细胞分裂的物体，经一次分裂而为二，二再分裂为四，四再分裂为八，如此继续分裂有限多次，而随机中止. 设分裂 n 次终止的概率为 $1/2^n (n=1, 2, \cdots)$，试求原物在分裂终止后所分成的子块数目的分布列.
4. 甲、乙两名篮球队员轮流投篮，直至某人投中为止. 如果甲投中的概率为 0.4，乙投中的概率为 0.6，试分别求甲、乙投篮次数的分布列.
5. 设 ξ 服从泊松分布，已知 $P(\xi=1)=2P(\xi=2)$，试求 $P(\xi=3)$.
6. 若 ξ 服从参数为 λ 的泊松分布，问 k 为何值时，$P(\xi=k)$ 最大？
7. 若 ξ 服从二项分布 $b(n, p)$，问 k 为何值时，$P(\xi=k)$ 最大？
8. 若 ξ 服从超几何分布(1)，问 k 为何值时，$P(\xi=k)$ 最大？
9. 某螺丝钉厂在生产稳定的条件下，废品率为 0.01，问一盒中应装多少只螺丝钉才能使其中含 100 只合格品的概率不小于 0.95？
10. 一本 500 页的书，共有 500 个错字，每个错字等可能地出现在每一页上，试求在给定的一页上至少有三个错字的概率.

第三节 连续型随机变量

3.1 连续型随机变量及其分布密度

若一随机变量 ξ 的分布函数可以表成某一非负函数 $p(x)$ 的积分：

$$F(x)=\int_{-\infty}^{x} p(t)dt, \quad -\infty<x<\infty, \tag{1}$$

则称ξ是一连续型随机变量，$p(x)$称为ξ的(概率)分布密度.

连续型随机变量也是在实际应用中经常见到的一种随机变量.

【例1】 考虑下面的几何型随机试验(参看第一章第四节)：向[0, 1]区间随机投一点，使所投的点落在[0, 1]内任一子区间内的概率与子区间的长度成正比，而与子区间所在的位置无关. 在这个随机试验中，设ξ是点落下后位置的坐标. ξ是一随机变量，它的分布函数

$$F(x)=P(\xi<x)$$

$$=\begin{cases}0, & \text{当 } x\leqslant 0;\\ \dfrac{[0,\ x]\text{的长度}}{[0,\ 1]\text{的长度}}=x, & \text{当 } 0<x\leqslant 1;\\ 1, & \text{当 } x>1.\end{cases} \tag{2}$$

这个分布函数可以表成下述函数

$$p(x)=\begin{cases}0, & \text{当 } x<0;\\ 1, & \text{当 } 0\leqslant x\leqslant 1;\\ 0, & \text{当 } x>1\end{cases} \tag{3}$$

的积分 $\qquad F(x)=\int_{-\infty}^{x} p(t)dt.$

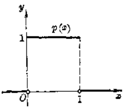


图 2-5

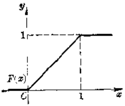


图 2-6

所以，这个 ξ 是一连续型随机变量，它的分布密度为(3)式定义的函数 $p(x)$。

对于连续型随机变量 ξ 来说，知道了它的分布密度 $p(x)$ 便可以知道它的分布函数 $F(x)$，所以我们将通过分布密度来掌握连续型随机变量的概率规律。

利用分布密度可以得出 ξ 落在任意区间内的概率：

$$P(a\leqslant\xi<b)=F(b)-F(a)=\int_a^b p(x)dx.$$

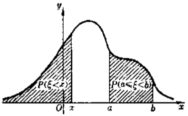


图 2-7

直观地来看，连续型随机变量在各区间内取值的概率等于在此区间上分布密度曲线下的曲边梯形的面积，参看图 2–7。

对任一常数 a,

$$P(\xi=a)=F(a+0)-F(a)=\lim_{b\to a+0}[F(b)-F(a)]$$
$$=\lim_{b\to a+0}\int_a^b p(x)dx=0,$$

所以，连续型随机变量取任一可能值的概率是 0. 由此可见，连续型随机变量与离散型随机变量不同，不能用列举它取各个可能值的概率来掌握它的概率规律，而必需通过它在各个区间取值的概率才能掌握它的概率规律。

容易看出，随机变量 ξ 的分布密度 $p(x)$ 具有下述基本性质：

1) $p(x)\geqslant 0,\quad -\infty<x<\infty;$

2) $\int_{-\infty}^{\infty}p(x)dx=1.$

反之，如果 $p(x)$ 满足 1), 2)，那么它一定是某一随机变

量的分布密度．这是因为，令

$$F(x)=\int_{-\infty}^{x} p(u)du,$$

可以验证，$F(x)$满足分布函数的三个基本性质1)～3)，因而$F(x)$是某一随机变量ξ的分布函数(第一节末尾的结论一)，$p(x)$就是这个ξ的分布密度．

还可以证明，ξ落在任一波雷尔集B内的概率，等于分布密度在这个集上的积分

$$P(\xi\in B)=\int_{B} p(x)dx. \tag{4}$$

证明这个结论需要利用测度与积分的知识，故从略．这里，我们结合特殊情形来说明它的含义．

$\{\xi<a\}$等价于"ξ落在区间$(-\infty, a)$内"，这就是说$\{\xi<a\}=\{\xi\in(-\infty, a)\}$，这时，公式(4)就是

$$P(\xi<a)=P(\xi\in(-\infty, a))=\int_{(-\infty, a)} p(x)dx.$$

这实际就是(1)的另一写法．

又如当$a\geqslant 0$时，$\{|\xi|>a\}$等价于$\{\xi\in B\}$，$B=\{x \mid |x|>a\}$，公式(4)即是

$$\begin{aligned}P(|\xi|>a)&=\int_{B} p(x)dx=\int_{|x|>a} p(x)dx\\&=\int_{-\infty}^{-a} p(x)dx+\int_{a}^{+\infty} p(x)dx.\end{aligned}$$

最后，我们指出，当ξ的分布函数$F(x)$已知时，如果$F(x)$连续，并且除去有限个点外，导函数$F'(x)$存在且连续，则$F(x)$可以表示为

$$F(x)=\int_{-\infty}^{x} F'(u)du,$$

因而ξ是连续型随机变量，它的分布密度就是$F'(x)$．今后

在求连续型随机变量的分布密度时，一般需按定义的要求来解决，在上述条件成立时，也可以使用求分布函数导函数的方法。

3.2 几个常见的连续型分布

1．均匀分布

向有限区间 $[a, b]$ 内随机投掷一个点的几何型随机试验中，设 ξ 是落点处的坐标，则 ξ 是一连续型随机变量，它的分布函数是

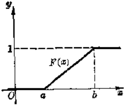


图 2-8

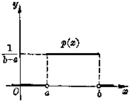


图 2-9

$$F(x)=\begin{cases}0, & \text{当 } x\leqslant a;\\ \dfrac{x-a}{b-a}, & \text{当 } a<x\leqslant b;\\ 1, & \text{当 } x>b.\end{cases} \tag{5}$$

分布密度为

$$p(x)=\begin{cases}0, & \text{当 } x<a;\\ \dfrac{1}{b-a}, & \text{当 } a\leqslant x\leqslant b;\\ 0, & \text{当 } x>b.\end{cases} \tag{6}$$

以(6)中 $p(x)$ 为分布密度的连续型分布称为 $[a, b]$ 上的均匀分布，简记作 $U[a, b]$．若随机变量 ξ 的分布为 $[a, b]$ 上的均匀分布，则称 ξ 具有均匀分布，或称 ξ 服从均匀分布，简记作

$$\xi \sim U[a, b].$$

本节例1中的随机变量具有[0, 1]上的均匀分布：$\xi \sim U[0, 1]$.

如上所述，均匀分布与直线上投点的几何型随机试验相联系，也就是说，若ξ在$[a, b]$中某区域内取值的概率与区域的长度成正比，而与区域的位置无关时，则ξ具有均匀分布.

2．指数分布

某些事物的"寿命"ξ（如电子元件的使用寿命，电话的通话时间，随机服务系统中的服务时间等），可以认为近似地具有下述性质：

$$P(\xi \geqslant t+s \mid \xi \geqslant t) = P(\xi \geqslant s), \quad t, s \geqslant 0, \tag{7}$$

即在寿命已达到t的条件下继续达到s以上的条件概率与从头开始算起寿命达到s以上的概率是一样的．现在我们来确定ξ的分布函数$F(t)=P(\xi<t)$，为此只需确定

$$G(t) = P(\xi \geqslant t) = 1 - F(t).$$

由于

$$P(\xi \geqslant t+s \mid \xi \geqslant t) = \frac{P(\xi \geqslant t+s,\ \xi \geqslant t)}{P(\xi \geqslant t)} = \frac{P(\xi \geqslant t+s)}{P(\xi \geqslant t)},$$

所以条件(7)等价于

$$G(t+s) = G(t)G(s), \quad t, s \geqslant 0.$$

利用第一章第九节的引理

$$G(t) = \theta^t, \quad t \geqslant 0.$$

由于$G(t)$是概率，$0<\theta<1$，可设$\theta = e^{-\lambda}$，$\lambda>0$，于是得

$$G(t) = e^{-\lambda t}, \qquad t \geqslant 0,$$

$$F(t) = 1 - e^{-\lambda t}, \quad t \geqslant 0,$$

而当$t<0$时，显然

$$F(t) = 0.$$

这样就得出我们所求的分布函数

$$F(t)=\begin{cases}1-e^{-\lambda t}, & t\geqslant 0;\\ 0, & t<0.\end{cases} \tag{8}$$

容易看出，ξ 具有分布密度

$$p(t)=\begin{cases}\lambda e^{-\lambda t}, & t\geqslant 0;\\ 0, & t<0.\end{cases} \tag{9}$$

以(9)中的 $p(t)$ 为分布密度的概率分布称为指数分布。指数分布含有一个参数 $\lambda>0$.

3．正态分布

以下列函数 $p(x)$ 为分布密度的分布称为正态分布

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}},\quad -\infty<x<\infty. \tag{10}$$

其中 a 和 σ 是常数且 $\sigma>0$. 正态分布含有两个参数 a, σ. 以 a, σ 为参数的正态分布简记为 $N(a,\sigma^2)$. 正态分布的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt,\quad -\infty<x<\infty. \tag{11}$$

以 $a=0$, $\sigma=1$ 为参数的正态分布 $N(0,1)$ 称为标准正态分布。相应的分布密度及分布函数通常记作 $\varphi(x)$ 及 $\Phi(x)$:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad -\infty<x<\infty; \tag{12}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt,\quad -\infty<x<\infty. \tag{13}$$

首先，我们来验证(10)中的 $p(x)$ 确能作为分布密度，即要验证 $p(x)$ 具有分布密度的基本性质 1)和 2)。性质 1) 显然成立，主要验证性质 2)。作变换 $\frac{x-a}{\sigma}=y$，有

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy. \tag{14}$$

而

$$\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\right)\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy\right)$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dx\,dy. \tag{15}$$

变换为极坐标：$x=r\cos\theta$, $y=r\sin\theta$，得

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dx\,dy=\frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi}e^{-\frac{r^2}{2}}r\,dr\,d\theta$$
$$=\int_0^{\infty}re^{-\frac{r^2}{2}}dr=1. \tag{16}$$

由(14)～(16)，并注意到 $\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$ 的非负性，得

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=1, \tag{17}$$

于是性质 2)得到证明.

正态分布 $N(a,\sigma^2)$ 的分布密度 $p(x)$ 在 $x=a$ 处达到极大，并且关于 $x=a$ 对称：

$$p(a-x)=p(a+x).$$

作出 $p(x)$ 的图象可以直观地看出，服从正态分布 $N(a,\sigma^2)$ 的随机变量经常在 a 的附近取值，而且在对 $x=a$ 对称的区间内取值的概率相同. 正态分布中的参数 a 表示相应随机变量取值的集中位置. 比较具有不同 σ 的正态分布密度，可以看到 σ 越小，$p(x)$ 在 $x=a$ 的两侧下降越快，表示相应随机变量取值越集中于 $x=a$ 附近；σ 越大，$p(x)$ 在 $x=a$ 两侧下降比较缓慢，表示相应的随机变量取值比较分散(参看图 2-10).

由于在自然界，工农业生产和科学实验中相当广泛的一类随机现象可以用正态分布或近似地用正态分布来刻划，所以正态分布是一个具有广泛应用的分布，它在随机变量的分布理论研究中占有十分重要的地位. 在以后的章节中我们将

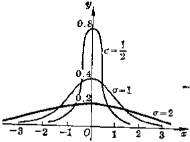


图 2-10

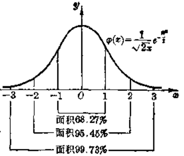


图 2-11

对正态分布做比较多的讨论.

正态分布密度和分布函数的数值，可以利用正态分布的数值表来计算. 本书附有标准正态分布函数 $\Phi(x)$ 的数值表，表中列出了对应于 $x \geqslant 0$ 的函数值. 利用对称性，可按下列公式求出对应于 $x<0$ 的相应函数值:

$$\varphi(-x)=\varphi(x), \quad \Phi(-x)=1-\Phi(x).$$

【例 2】 设 ξ 服从标准正态分布: $\xi \sim N(0, 1)$，试求 $P(|\xi|<1)$, $P(|\xi|<2)$, $P(|\xi|<3)$.

解:

$$P(|\xi|<1)=\Phi(1)-\Phi(-1)=\Phi(1)-(1-\Phi(1))$$
$$=2\Phi(1)-1 \doteq 68.27\%,$$
$$P(|\xi|<2)=2\Phi(2)-1 \doteq 95.45\%,$$
$$P(|\xi|<3)=2\Phi(3)-1 \doteq 99.73\%.$$

从附表 2 查得

$$\Phi(1)=0.841345, \quad \Phi(2)=0.977250,$$
$$\Phi(3)=0.998650,$$

代入式中计算即有所得结果.

【例 3】 已知 $\xi \sim N(0, 1)$，求 x 使 $P(|\xi|>x)=0.10$,

0.05, 0.01.

解:

$$P(|\xi|>x)=2P(\xi<-x)=2(1-\Phi(x)),$$

$$\Phi(x)=1-\frac{1}{2}P(|\xi|>x).$$

求 x 使 $P(|\xi|>x)=0.1$, 相当于求 x 使 $\Phi(x)=0.95$. 从表上可查得 $x\approx1.65$, 即

$$P(|\xi|>1.65)=0.10.$$

类似地可得

$$P(|\xi|>1.96)=0.05,\ P(|\xi|>2.58)=0.01.$$

为了求出正态分布 $N(a,\sigma^2)$ 的分布函数 $F(x)$ 的值, 可以利用下列变换公式

$$\begin{aligned}F(x)&=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(t-a)^2}{2\sigma^2}}dt\\&=\int_{-\infty}^{\frac{x-a}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du\quad\left(u=\frac{t-a}{\sigma}\right)\\&=\Phi\left(\frac{x-a}{\sigma}\right).\end{aligned}\tag{18}$$

【例 4】 设 $\xi\sim N(2.3,\ 4)$, 则

$$\begin{aligned}P(2<\xi<4)&=F(4)-F(2)\\&=\Phi\left(\frac{4-2.3}{2}\right)-\Phi\left(\frac{2-2.3}{2}\right)\\&=\Phi(0.85)-\Phi(-0.15)\\&=\Phi(0.85)-(1-\Phi(0.15))\\&=0.3619.\end{aligned}$$

利用变换公式(18)及例 2 的结果, 可知对服从 $N(a,\sigma)$ 的随机变量 ξ 来说

$$P(|\xi-a|<\sigma)=68.27\%,$$

$$P(|\xi-a|<2\sigma)=95.45\%,$$
$$P(|\xi-a|<3\sigma)=99.73\%.$$

这使我们对前面所说服从正态分布的随机变量经常在 $x=a$ 的附近取值，以及 σ 的大小反映相应随机变量取值分散和集中的含义有更加确切的了解.

在前节和本节中，我们讨论了离散型和连续型随机变量，它们是实际问题中经常用到的两类随机变量. 应该指出，除上述两类随机变量外，还存在既非离散型又非连续型的随机变量. 在本书中我们将主要讨论这两类随机变量.

习 题 三

1. 已知 ξ 在 $[2, 4]$ 上均匀分布，试求
   $P(-1\leqslant\xi<3)$, $P((\xi-3)^2<0.25)$, $P(3\xi+2<11.6)$.
2. 已知 $\xi\sim N(2.5, 4)$，试求
   $P(\xi>5)$, $P(\xi<-1)$, $P(|\xi-2|<3)$, $P(2\xi^2+1>5)$.
3. 已知 $\xi\sim N(0, 1)$，试求 x_1, x_2, x_3，使落在 $(-\infty, x_1)$，(x_1, x_2)，(x_2, x_3)，$(x_3, +\infty)$ 的概率之比为 5:4:3:6.
4. 测量到某一目标的距离时发生的随机误差 ξ(米)，具有分布密度
   $$p(x)=\frac{1}{40\sqrt{2\pi}}e^{-\frac{(x-20)^2}{3200}}.$$
   求在三次测量中至少有一次误差的绝对值不超过 30 米的概率.
5. 设某类电子管的寿命具有分布密度
   $$f(x)=\begin{cases}\dfrac{100}{x^2}, & x>100;\\ 0, & x\leqslant 100.\end{cases}$$
   一架无线电在开初使用的 150 小时中，三个这类电子管没有一个要替换的概率是什么? 在开初使用的 150 小时中三个这类电子管全部要替换的概率又是什么?
6. 设随机变量 ξ 的分布函数为

$$F(x)=\begin{cases}0, & x<0;\\ Cx^2, & 0\leqslant x<1;\\ 1, & x\geqslant 1.\end{cases}$$

试求(1)常数 C; (2)ξ 落在(0.3, 0.7)内的概率; (3)ξ 的分布密度.

7. 设随机变量 ξ 的分布函数为

$$F(x)=A+B\operatorname{arctg}x,$$

试求(1)常数 A 与 B; (2)ξ 的分布密度; (3)$P(-1<\xi^3<8)$.

8. 已知随机变量 ξ 的分布密度为

$$f(x)=\begin{cases}Cx^3, & 0<x<1;\\ 0, & \text{其它}.\end{cases}$$

(1)确定常数 C; (2)求数 a, 使 $P(\xi>a)=P(\xi<a)$; (3)求数 b, 使 $P(\xi>b)=0.01$.

9. 设随机变量 ξ 的分布密度为

$$f(x)=\begin{cases}Ax^2e^{-kx}, & 0\leqslant x<\infty;\\ 0, & x<0.\end{cases}$$

其中 $k>0$ 是给定的常数. 试求(1)未知常数 A; (2)$P\left(0<\xi<\frac{1}{k}\right)$; (3)$\xi$ 的分布函数.

10. 设 ξ 在(0, 3)上均匀分布, 求方程 $4x^2+4\xi x+\xi+2=0$ 有实根的概率.

第四节 随机向量

4.1 随机向量及联合分布

在许多随机试验中, 讨论的问题往往不只涉及一个随机变量, 而是需要同时考虑若干个随机变量. 例如, 在会面问题中(第一章第四节例 1), 需要考虑两个随机变量: 甲到达的时刻 ξ 和乙到达的时刻 η, 我们所关心的是它们构成的点 (ξ, η) 的分布规律.

在本节中, 我们讨论由概率空间$(\Omega, \mathscr{F}, P)$上的 n 个随

机变量 $\xi_1, \xi_2, \cdots, \xi_n$ 组成的向量(或 n 维空间的点) $\xi=(\xi_1, \xi_2, \cdots, \xi_n)$，这个向量实际上表示 Ω 上的一个向量值的函数：对应于每一样本点 ω，它取一向量值

$$\xi(\omega)=(\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega)).$$

在每次试验中，随着不同样本点的出现，ξ 随机地落在 n 维空间中的不同位置，我们所关心的是"ξ 落在某一区域内"这种事件的概率．我们所要讨论的事件用下面的记号来表示：

$$\begin{aligned}&\{\xi_1<x_1, \xi_2<x_2, \cdots, \xi_n<x_n\}\\&=\{\xi_1<x_1\}\cap\{\xi_2<x_2\}\cap\cdots\cap\{\xi_n<x_n\}\\&=\{\omega \mid \xi_1(\omega)<x_1, \xi_2(\omega)<x_2, \cdots, \xi_n(\omega)<x_n\}.\end{aligned}$$

这表示"$\xi=(\xi_1, \xi_2, \cdots, \xi_n)$落在 n 维区间$(-\infty, x_1)\times(-\infty, x_2)\times\cdots\times(-\infty, x_n)$ *) 内"这一事件，$n=2$ 的情形参看图 2-12．一般，"$\xi=(\xi_1, \xi_2, \cdots, \xi_n)$落在区域 B 内"这一事件记作

$$\begin{aligned}&\{(\xi_1, \xi_2, \cdots, \xi_n)\in B\}\\&=\{\omega \mid (\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega))\in B\}.\end{aligned}$$

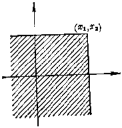


图 2-12

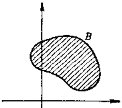


图 2-13

定义 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是同一概率空间 $(\Omega, \mathscr{F}, P)$ 上的 n 个随机变量，则称 Ω 上的向量函数

*) $(-\infty, x_1)\times\cdots\times(-\infty, x_n)=\{(y_1, \cdots, y_n) \mid -\infty<y_1<x_1, \cdots, -\infty<y_n<x_n\}$.

$$\xi(\omega)=(\xi_1(\omega),\ \xi_2(\omega),\ \cdots,\ \xi_n(\omega)),\quad \omega\in\Omega \tag{1}$$

为$(\Omega,\ \mathscr{F},\ P)$上的一个$n$维随机向量或$n$维随机变量．我们经常将随机向量(1)简记作$\xi=(\xi_1,\ \xi_2,\ \cdots,\ \xi_n)$．对于任意实数$x_1,\ x_2,\ \cdots,\ x_n$，概率

$$F(x_1,\ x_2,\ \cdots,\ x_n)=P(\xi_1<x_1,\ \xi_2<x_2,\ \cdots,\ \xi_n<x_n)$$

是$x_1,\ x_2,\ \cdots,\ x_n$的函数，称为随机向量$\xi=(\xi_1,\ \xi_2,\ \cdots,\ \xi_n)$的(联合)分布函数．

理论上可以证明，由随机向量ξ的分布函数$F(x_1,\ \cdots,\ x_n)$，可以确定出ξ落在任意n维波雷尔集B内的概率

$$P((\xi_1,\ \xi_2,\ \cdots,\ \xi_n)\in B),$$

所以，知道了分布函数便认为对随机向量的概率规律有了全面的了解．在概率论中研究随机向量基本要求之一就是掌握它的分布函数．

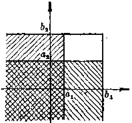


图 2-14

利用分布函数可以得出事件$\{a_1\leqslant\xi_1<b_1,\ a_2\leqslant\xi_2<b_2,\ \cdots,\ a_n\leqslant\xi_n<b\}$的概率．以$n=2$的情形为例，有

$$P(a_1\leqslant\xi_1<b_1,\ a_2\leqslant\xi_2<b_2)$$
$$=F(b_1,\ b_2)-F(a_1,\ b_2)-F(b_1,\ a_2)+F(a_1,\ a_2). \tag{2}$$

这个公式可从图 2-14 看出．它的推导如下：

$$P(a_1\leqslant\xi_1<b_1,\ a_2\leqslant\xi_2<b_2)$$
$$=P(a_1\leqslant\xi_1<b_1,\ \xi_2<b_2)-P(a_1\leqslant\xi_1<b_1,\ \xi_2<a_2)$$
$$=P(\xi_1<b_1,\ \xi_2<b_2)-P(\xi_1<a_1,\ \xi_2<b_2)$$
$$-P(\xi_1<b_1,\ \xi_2<a_2)+P(\xi_1<a_1,\ \xi_2<a_2).$$

一般，可用数学归纳法证明：

$$
\begin{aligned}
&P(a_1\leqslant\xi_1<b_1,\ \cdots,\ a_n\leqslant\xi_n<b_n)\\
&=F(b_1,\ \cdots,\ b_n)-\sum_{1\leqslant i\leqslant n}F(\cdots,\ a_i,\ \cdots)\\
&\quad+\sum_{1\leqslant i_1<i_2\leqslant n}F(\cdots,\ a_{i_1},\ \cdots,\ a_{i_2},\ \cdots)\\
&\quad+\cdots+(-1)^k\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}F(\cdots,\ a_{i_1},\ \cdots,\ a_{i_k},\ \cdots)\\
&\quad+\cdots+(-1)^nF(a_1,\ \cdots,\ a_n),
\end{aligned}
\tag{3}
$$

其中 $F(\cdots,\ a_{i_1},\ \cdots,\ a_{i_k},\ \cdots)$ 表示第 $i_1,\ \cdots,\ i_k$ 自变量分别为 $a_{i_1},\ \cdots,\ a_{i_k}$，而其余自变量 x_j 为 b_j 时 F 的函数值.

分布函数具有下列基本性质：

1）对于每一变量 x_i 来说，$F(x_1,\ \cdots,\ x_n)$是一不降函数；

2）$F(x_1,\ \cdots,\ \overset{(i)}{-\infty},\ \cdots,\ x_n)=0,\quad i=1,\ 2,\ \cdots,\ n,$

$$F(+\infty,\ \cdots,\ +\infty)=1,$$

其中

$$F(x_1,\ \cdots,\ \overset{(i)}{-\infty},\ \cdots,\ x_n)=\lim_{x_i\to-\infty}F(x_1,\ \cdots,\ x_i,\ \cdots,\ x_n),$$

$$F(+\infty,\ \cdots,\ +\infty)=\lim_{\substack{x_1\to+\infty\\ \cdots\\ x_n\to+\infty}}F(x_1,\ \cdots,\ x_n);$$

3）对于每一变量 x_i 来说，$F(x_1,\ \cdots,\ x_n)$左连续；

4）若 $a_i\leqslant b_i$，$i=1,\ \cdots,\ n$，则(3)式的右端非负.

以上各性质的证明如下：

1）可由下列事件的包含关系得出

$$\{\cdots,\ \xi_i<a_i,\ \cdots\}\subseteq\{\cdots,\ \xi_i<b_i,\ \cdots\},\quad 当\ a_i<b_i;$$

2)的第一式，可由下列关系式得出

$$
\begin{aligned}
0\leqslant F(x_1,\ \cdots,\ x_n)&=P(\xi_1<x_1,\ \cdots,\ \xi_n<x_n)\leqslant P(\xi_i<x_i)\\
&=F_{\xi_i}(x_i),
\end{aligned}
$$

其中 $F_{\xi_i}(x_i)$表示 ξ_i 的分布函数；

2)的第二式，可由下列关系式导出：

$$\overline{\{\xi_1<x_1,\ \cdots,\ \xi_n<x_n\}}=\overline{\{\xi_1<x_1\}}\cup\cdots\cup\overline{\{\xi_n<x_n\}},$$

$$1-F(x_1,\ \cdots,\ x_n)\leqslant[1-F_{\xi_1}(x_1)]+\cdots+[1-F_{\xi_n}(x_n)],$$

$$1-[1-F_{\xi_1}(x_1)]-\cdots-[1-F_{\xi_n}(x_n)]$$

$$\leqslant F(x_1,\ \cdots,\ x_n)\leqslant 1;$$

3) 与一维情形类似地证明;

4) 可由(3)式立刻看出.

经常讨论的随机向量有两种类型:离散型和连续型.

若随机向量 $\xi=(\xi_1,\ \cdots,\ \xi_n)$ 只取有限个或可数多个不同的向量值,则称 ξ 是一**离散型随机向量**. 设 ξ 的全部可能值为

$$(x_{1i_1},\ x_{2i_2},\ \cdots,\ x_{ni_n}),\quad i_1,\ i_2,\ \cdots,\ i_n=1,\ 2,\ \cdots,$$

则 ξ 取各个可能值的概率

$$p_{i_1,i_2,\cdots,i_n}=P(\xi_1=x_{1i_1},\ \xi_2=x_{2i_2},\ \cdots,\ \xi_n=x_{ni_n}),$$
$$i_1,\ i_2,\ \cdots,\ i_n=1,\ 2,\ \cdots,$$

称为 ξ 的**分布列**. 对于一个离散型的随机向量,知道了分布列便可以得出它的分布函数,以及 ξ 落在任何区域内的概率. 所以,我们将通过分布列来掌握离散型随机向量的分布规律.

【例 1】 设有五件产品,其中两件是次品,三件是正品. 从五件产品中任意依次取出两件,每次取出后不放回. 考虑第一次取到的次品件数 ξ_1, 第二次取到的次品件数 ξ_2. 由于它们分别取 0、1 两个值,因此,$(\xi_1,\ \xi_2)$ 可能取得的向量值 (0,0), (1, 0), (0, 1), (1, 1)是一离散型二维随机向量,它的分布列为

$$P(\xi_1=0,\ \xi_2=0)=\frac{C_3^1C_2^1}{C_5^1C_4^1}=\frac{6}{20},$$

$$P(\xi_1=1,\ \xi_2=0)=\frac{C_2^1C_3^1}{C_5^1C_4^1}=\frac{6}{20},$$

$$P(\xi_1=0,\ \xi_2=1)=\frac{C_3^1C_2^1}{C_5^1C_4^1}=\frac{6}{20},$$

$$P(\xi_1=1,\ \xi_2=1)=\frac{C_2^1C_1^1}{C_5^1C_4^1}=\frac{2}{20}.$$

【例 2】 在例 1 的问题中，如果每次取出后放回，则 (ξ_1,ξ_2) 的分布列为

$$P(\xi_1=0,\ \xi_2=0)=\frac{C_3^1C_3^1}{5^2}=\frac{9}{25},$$

$$P(\xi_1=1,\ \xi_2=0)=\frac{C_2^1C_3^1}{5^2}=\frac{6}{25},$$

$$P(\xi_1=0,\ \xi_2=1)=\frac{C_3^1C_2^1}{5^2}=\frac{6}{25},$$

$$P(\xi_1=1,\ \xi_2=1)=\frac{C_2^1C_2^1}{5^2}=\frac{4}{25}.$$

现在介绍一个重要的离散型分布.

多项分布 设独立地进行了 n 次试验，在每次试验中，有 k 个互不相容的事件 $A_1,\ \cdots,\ A_k$ 之一出现，并且

$$P(A_i)=p_i,\quad i=1,\ \cdots,\ k,$$

$$p_1+\cdots+p_k=1$$

在每次试验中不变. 令 ξ_i 是在 n 次独立试验中 A_i 出现的次数，则 $(\xi_1,\ \cdots,\ \xi_{k-1})$ 是一个 $k-1$ 维随机变量，它的分布列为

$$P(\xi_1=n_1,\ \cdots,\ \xi_{k-1}=n_{k-1})$$

$$=\begin{cases}\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}}\\ \quad\times p_1^{n_1}\cdots p_{k-1}^{n_{k-1}}p_k^{n-n_1-\cdots-n_{k-1}},\\ \quad\quad \text{当 } n_1,\ \cdots,\ n_{k-1} \text{ 为非负整数, } n_1+\cdots+n_{k-1}\leqslant n;\\ 0,\quad \text{其它情形.}\end{cases}$$

或

$$P(\xi_1=n_1, \cdots, \xi_{k-1}=n_{k-1})$$
$$=\begin{cases} \dfrac{n!}{n_1!\cdots n_k!} p_1^{n_1}\cdots p_k^{n_k}, & \text{当 } n_1, \cdots, n_k \text{ 为非负整数,} \\ & n_1+\cdots+n_k=n; \\ 0, & \text{其它情形.} \end{cases} \quad (4)$$

(4)中的分布列称为参数为$(n; p_1, \cdots, p_{k-1})$的多项分布. 当随机向量$(\xi_1, \cdots, \xi_{k-1})$有分布列(4)时, 称为服从参数为$(n; p_1, \cdots, p_{k-1})$的多项分布.

多项分布是二项分布的推广. 容易看出, 当$k=2$时,多项分布就是二项分布.

如果随机向量$\xi=(\xi_1, \xi_2, \cdots, \xi_n)$的分布函数可以表成某一非负函数$p(x_1, x_2, \cdots, x_n)$的积分

$$F(x_1, x_2, \cdots, x_n)=P(\xi_1<x_1, \xi_2<x_2, \cdots, \xi_n<x_n)$$
$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} p(t_1, t_2, \cdots, t_n)dt_1dt_2\cdots dt_n,$$

则ξ称为连续型随机向量, $p(x_1, x_2, \cdots, x_n)$称为ξ的分布密度.

例如, 在向n维空间某一区域G内等可能地投点的几何型随机试验中,令$(\xi_1, \xi_2, \cdots, \xi_n)$是落点处的坐标,则$(\xi_1, \xi_2, \cdots, \xi_n)$落在$G$内任意可求积区域$B$内的概率为

$$P((\xi_1, \cdots, \xi_n)\in B)=\frac{[B\text{ 的度量}]}{[G\text{ 的度量}]}$$
$$=\int\cdots\int_B p(x_1, \cdots, x_n)dx_1\cdots dx_n, \quad (5)$$

其中

$$p(x_1, \cdots, x_n)=\begin{cases} \dfrac{1}{S}, & \text{当 } (x_1, \cdots, x_n)\in G, \\ 0, & \text{其它} (x_1, \cdots, x_n); \end{cases} \quad (6)$$
$$S=[G\text{ 的度量}].$$

所以，$(\xi_1, \cdots, \xi_n)$是一个连续型随机向量，它的分布密度为(6)中的$p(x_1, \cdots, x_n)$. (5)中的分布称为区域G上的均匀分布.

和一维随机变量的情形相同，分布密度具有下列基本性质：

1) $p(x_1, \cdots, x_n) \geqslant 0, \quad -\infty < x_i < \infty, \ i=1, \cdots, n;$

2) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1.$

利用测度和积分的理论可以证明，对于任意n维波雷尔集B,

$$P((\xi_1, \cdots, \xi_n) \in B) = \int \cdots \int_B p(x_1, \cdots, x_n) dx_1 \cdots dx_n. \tag{7}$$

反之，若n元函数$p(x_1, \cdots, x_n)$具有1), 2)两条性质，它一定是某一n维随机向量的分布密度.

另一个重要的多元连续型分布是

多元正态分布 设$B=(b_{ij})$是一n阶正定(对称)方阵，$B^{-1}=(r_{ij})$是它的逆矩阵，$|B|$表示B的行列式，$a=(a_1, a_2, \cdots, a_n)'$是一个实值列向量(这里“$'$”表示矩阵的转置). 以函数

$$p(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{n} r_{ij}(x_i - a_i)(x_j - a_j)\right\} \tag{8}$$

为分布密度的概率分布称为具有参数(a, B)的n元正态分布，简记作$N_n(a, B)$或$N(a, B)$.

正态分布的分布密度(8)也常常用向量的形式表示成

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-a)'B^{-1}(x-a)\right\}, \tag{9}$$

其中 $x=(x_1, \cdots, x_n)'$.

$n=1$ 时, 同我们以前讲的(一元)正态分布完全一致.

$n=2$ 时, 正态分布的分布密度常常表示成另一种具体的形式. 设 $B=\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. 由于 B 是正定阵, 因此, $b_{11}>0$, $b_{22}>0$, $b_{12}^2<b_{11}b_{22}$. 记 $b_{11}=\sigma_1^2$, $b_{22}=\sigma_2^2$, 则 $b_{12}=b_{21}=r\sigma_1\sigma_2$, 其中 $\sigma_1>0$, $\sigma_2>0$, $|r|<1$. 这就是说, 任意 2×2 正定阵 B 都可以表成

$$B=\begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

因而二元正态分布的分布密度有如下形式

$$p(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}. \tag{10}$$

4.2 边沿分布

对于给定的随机向量 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)$, 我们也常常需要知道它的某 k 个分量构成的随机向量 $(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k})$ 的分布规律. 通常称

$$F_{\xi_1}(x_1)=P(\xi_1<x_1), \cdots, F_{\xi_n}(x_n)=P(\xi_n<x_n)$$

为 ξ 的(一维)边沿分布函数; 称

$$F_{\xi_1,\xi_2}(x_1, x_2)=P(\xi_1<x_1, \xi_2<x_2), \cdots,$$
$$F_{\xi_i,\xi_j}(x_i, x_j)=P(\xi_i<x_i, \xi_j<x_j), \cdots$$

为 ξ 的(二维)边沿分布函数; 一般, 称

$$F_{\xi_{i_1},\cdots,\xi_{i_k}}(x_{i_1}, \cdots, x_{i_k})=P(\xi_{i_1}<x_{i_1}, \cdots, \xi_{i_k}<x_{i_k})$$

为ξ的(k维)边沿分布函数.

边沿分布可以由联合分布来确定，下面我们为了叙述简明，只讨论$n=2$的情形. 所得结果都可以推广到一般情形.

设(ξ, η)的分布函数为$F(x, y)$，现在要由此求出ξ的分布函数$F_\xi(x)$及η的分布函数$F_\eta(y)$.

由于
$$F_\xi(x)=P(\xi<x),$$
注意到事件的关系：
$$\{\xi<x\}=\{\xi<x,\ \eta<1\}\cup\{\xi<x,\ \eta<2\}\cup\cdots\cup\{\xi<x,\ \eta<n\}\cup\cdots,$$
$$\{\xi<x,\ \eta<1\}\subseteq\{\xi<x,\ \eta<2\}\subseteq\cdots\subseteq\{\xi<x,\ \eta<n\}\subseteq\cdots.$$
利用概率的下连续性
$$F_\xi(x)=\lim_{n\to+\infty}P(\xi<x,\ \eta<n)=\lim_{n\to+\infty}F(x,\ n).$$
又由于$F(x, y)$是y的下降函数,
$$\lim_{n\to+\infty}F(x,\ n)=\lim_{y\to+\infty}F(x,\ y)=F(x,\ +\infty).$$
这样我们就得到关于边沿分布的公式
$$F_\xi(x)=F(x,\ +\infty), \tag{11}$$
$$F_\eta(y)=F(+\infty,\ y), \tag{12}$$
其中$F(x,\ +\infty)=\lim\limits_{y\to+\infty}F(x,\ y)$，$F(+\infty,\ y)=\lim\limits_{x\to+\infty}F(x,\ y)$.

下面讨论离散型的情形.

设(ξ, η)是离散型随机向量，它的分布列为
$$p_{ij}=P(\xi=x_i,\ \eta=y_j),\quad i,\ j=1,\ 2,\ \cdots.$$
显然，ξ所取的一切可能值为x_i, $i=1, 2, \cdots$. 由于事件$\{\xi=x_i\}$可以表成互不相容事件之和：$\{\xi=x_i\}=\bigcup\limits_{j=1}^{\infty}\{\xi=x_i, \eta=y_j\}$，所以，$\xi$的分布列为

$$P(\xi=x_i)=\sum_{j=1}^{\infty}p_{ij}. \tag{13}$$

类似地，η 的分布列为

$$P(\eta=y_j)=\sum_{i=1}^{\infty}p_{ij}. \tag{14}$$

【例 3】 考虑本节例 1 中的(ξ_1, ξ_2)，在那里已经求出了它的联合分布，利用公式(13)、(14)又可得出边沿分布. ξ_1 的分布列为

$$P(\xi_1=0)=P(\xi_1=0,\ \xi_2=0)+P(\xi_1=0,\ \xi_2=1)$$
$$=\frac{6}{20}+\frac{6}{20}=\frac{3}{5},$$
$$P(\xi_1=1)=P(\xi_1=1,\ \xi_2=0)+P(\xi_1=1,\ \xi_2=1)$$
$$=\frac{6}{20}+\frac{2}{20}=\frac{2}{5}.$$

ξ_2 的分布列为

$$P(\xi_2=0)=P(\xi_1=0,\ \xi_2=0)+P(\xi_1=1,\ \xi_2=0)$$
$$=\frac{6}{20}+\frac{6}{20}=\frac{3}{5},$$
$$P(\xi_2=1)=P(\xi_1=0,\ \xi_2=1)+P(\xi_1=1,\ \xi_2=1)$$
$$=\frac{6}{20}+\frac{2}{20}=\frac{2}{5}.$$

如果将(ξ_1, ξ_2)的联合分布列用一矩形表表示(见表 2-2)，则 ξ_1 的分布列可通过表中各列相加而得到，ξ_2 的分布列可通过表中各行相加而得到.

类似地，本节例 2 中的(ξ_1, ξ_2)的联合分布及边沿分布可列成表 2-3.

以上两个随机向量的边沿分布相同，但它们的联合分布却不相同，这说明仅仅知道了边沿分布，并不能确定出联合分布.

表 2-2

ξ_2	ξ_1		
	0	1	$P(\xi_2=y_j)$
0	$\frac{6}{20}$	$\frac{6}{20}$	$\frac{3}{5}$
1	$\frac{6}{20}$	$\frac{2}{20}$	$\frac{2}{5}$
$P(\xi_1=x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

表 2-3

ξ_2	ξ_1		
	0	1	$P(\xi_2=y_j)$
0	$\frac{9}{25}$	$\frac{6}{25}$	$\frac{3}{5}$
1	$\frac{6}{25}$	$\frac{4}{25}$	$\frac{2}{5}$
$P(\xi_1=x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

现在讨论(ξ, η)是连续型随机向量的情形.

设(ξ, η)有分布密度 $p(x, y)$，利用(11)及(12)可知，ξ 和 η 的分布函数分别为

$$F_\xi(x)=F(x, +\infty)=\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty}p(t, s)ds\right)dt,$$

$$F_\eta(y)=F(+\infty, y)=\int_{-\infty}^{y}\left(\int_{-\infty}^{\infty}p(t, s)dt\right)ds.$$

这样,我们就得出结论:

连续型随机向量(ξ, η)的边沿分布仍为连续型,并且 ξ 和 η 的分布密度 $p_\xi(x)$和 $p_\eta(y)$可用联合分布密度表出:

$$p_\xi(x)=\int_{-\infty}^{\infty}p(x,\ y)dy, \tag{15}$$

$$p_\eta(y)=\int_{-\infty}^{\infty}p(x,\ y)dx. \tag{16}$$

【例 4】 设 $(\xi,\ \eta)$ 服从二元正态分布，它的分布密度为(参看(10))

$$p(x,\ y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\times\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\},$$

利用(15)，(16)立即可以得出 ξ 和 η 的分布密度．在下面计算中，令 $\frac{x-a}{\sigma_1}=u,\ \frac{y-b}{\sigma_2}=v.$

$$\begin{aligned}
p_\xi(x)&=\int_{-\infty}^{\infty}p(x,\ y)dy\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\times\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}[u^2-2ruv+v^2]\right\}dv\\
&=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{u^2}{2}}\times\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\exp\left\{-\frac{r^2u^2-2ruv+v^2}{2(1-r^2)}\right\}dv\\
&=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{u^2}{2}}\times\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}e^{-(v-ru)^2/2(1-r^2)}dv\\
&=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{u^2}{2}}=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{(x-a)^2}{2\sigma_1^2}}.
\end{aligned}$$

同理 $$p_\eta(y)=\frac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(y-b)^2}{2\sigma_2^2}}.$$

由此看出：二元正态分布的边沿分布仍为正态分布.

一般情形有类似结论，证明也类似，只是表述形式比较复杂一些.

若$(\xi_1, \cdots, \xi_n)$的分布函数为$F(x_1, \cdots, x_n)$，则它的k个分量$(\xi_{i_1}, \cdots, \xi_{i_k})$$(1\leqslant i_1<\cdots<i_k\leqslant n, 1\leqslant k\leqslant n)$的分布函数为

$$F_{\xi_{i_1},\cdots,\xi_{i_k}}(x_{i_1}, \cdots, x_{i_k}) = \lim_{\substack{x_j\to+\infty \\ j\neq i_1,\cdots,i_k}} F(x_1, \cdots, x_n)$$

$$= F(c_1, c_2, \cdots, c_n), \tag{17}$$

其中$c_{i_r}=x_{i_r}(r=1, 2, \cdots, n)$，其余$c_i$为$+\infty$.

若离散型随机向量$(\xi_1, \cdots, \xi_n)$的分布列为

$$p_{j_1,\cdots,j_n}=P(\xi_1=x_{1j_1}, \cdots, \xi_n=x_{nj_n}),$$

$$j_1, \cdots, j_n=1, 2, \cdots,$$

则$(\xi_{i_1}, \cdots, \xi_{i_k})$的分布列为

$$P(\xi_{i_1}=x_{i_1j_{i_1}}, \cdots, \xi_{i_k}=x_{i_kj_{i_k}}) = \sum_{j_1=b_1}^{c_1}\cdots\sum_{j_n=b_n}^{c_n} p_{j_1,\cdots,j_n},$$

其中$b_{i_r}=c_{i_r}=j_{i_r}(r=1, 2, \cdots, k)$，其余$b_i=1$, $c_i=+\infty$.

若连续型随机向量$(\xi_1, \cdots, \xi_n)$的分布密度为$p(x_1, \cdots, x_n)$，则$(\xi_{i_1}, \cdots, \xi_{i_k})$有分布密度

$$p_{\xi_{i_1},\cdots,\xi_{i_k}}(x_{i_1}, \cdots, x_{i_k}) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p(x_1, \cdots, x_n)dx_{j_1}\cdots dx_{j_{n-k}}, \tag{18}$$

集合$\{j_1, \cdots, j_{n-k}\}=\{1, 2, \cdots, n\}-\{i_1, \cdots, i_k\}$.

4.3 条件分布

当同时考察若干个随机变量时，常常需要考虑：已知某些随机变量取得某值的条件下，另一些随机变量取值的条件概率.

我们对二维的情形来讨论.

先考虑离散型的随机向量(ξ, η)，设它的联合分布列为

$p_{ij}=P(\xi=x_i,\ \eta=y_j)$. 在已知 $\eta=y_j$ 的条件下(设 $P(\eta=y_j)>0$)ξ 取各可能值的条件概率

$$P(\xi=x_i|\eta=y_j)=\frac{P(\xi=x_i,\ \eta=y_j)}{P(\eta=y_j)}=\frac{p_{ij}}{\sum_{i=1}^{\infty}p_{ij}},$$

$$i=1,\ 2,\ \cdots, \tag{19}$$

称为在 $\eta=y_j$ 的条件下 ξ 的条件分布.

同样,当 $P(\xi=x_i)>0$ 时,

$$P(\eta=y_j|\xi=x_i)=\frac{P(\xi=x_i,\ \eta=y_j)}{P(\xi=x_i)}=\frac{p_{ij}}{\sum_{j=1}^{\infty}p_{ij}},$$

$$j=1,\ 2,\ \cdots, \tag{20}$$

称为在 $\xi=x_i$ 的条件下 η 的条件分布.

对于一般的随机向量$(\xi,\ \eta)$,由于可能出现 $P(\eta=y)=0$,因此不能直接用条件概率的公式来定义条件分布函数. 为了在一般情况下定义条件分布函数,一种自然的想法是先考虑

$$P(\xi<x|y\leqslant\eta<y+\Delta y)=\frac{P(\xi<x,\ y\leqslant\eta<y+\Delta y)}{P(y\leqslant\eta<y+\Delta y)},$$

然后令 $\Delta y\to 0$, 将上式的极限定义为 $\eta=y$ 之下 ξ 的条件分布.

详细地说,设 y 是取定的常数,且对任意 $\Delta y>0$, $P(y\leqslant\eta<y+\Delta y)>0$, 如果极限

$$\lim_{\Delta y\to 0}P(\xi<x|y\leqslant\eta<y+\Delta y)$$

存在,则称此极限为在 $\eta=y$ 的条件下 ξ 的条件分布函数,并用记号 $P(\xi<x|\eta=y)$ 表示,或简记为 $F_{\xi|\eta}(x|y)$.

现在讨论$(\xi,\ \eta)$是连续型随机向量的情形. 设它的联合分布密度为 $p(x,\ y)$, ξ, η 的分布密度分别为 $p_\xi(x)$, $p_\eta(y)$. 并且假定 $p(x,\ y)$, $p_\xi(x)$, $p_\eta(y)$ 是连续函数. 注意到在上述

假定下，对任意 x，

$$g(v)=\int_{-\infty}^{x}p(u,\ v)du$$

是 v 的连续函数．利用中值定理，当 $p_{\eta}(y)>0$ 时，

$$\lim_{\Delta y\to 0}P(\xi<x|y\leqslant\eta<y+\Delta y)$$

$$=\lim_{\Delta y\to 0}\frac{\int_{y}^{y+\Delta y}\left(\int_{-\infty}^{x}p(u,\ v)du\right)dv}{\int_{y}^{y+\Delta y}p_{\eta}(v)dv}$$

$$=\lim_{\Delta y\to 0}\frac{\int_{-\infty}^{x}p(u,\ y+\theta_1\Delta y)du}{p_{\eta}(y+\theta_2\Delta y)}\quad(|\theta_1|<1,\ |\theta_2|<1)$$

$$=\frac{\int_{-\infty}^{x}p(u,\ y)du}{p_{\eta}(y)},$$

所以，

$$F_{\xi|\eta}(x|y)=P(\xi<x|\eta=y)=\int_{-\infty}^{x}\frac{p(u,\ y)}{p_{\eta}(y)}du.$$

由此可见，在 $\eta=y$ 的条件下 ξ 的条件分布函数为连续型，条件分布密度为

$$p_{\xi|\eta}(x|y)=\frac{p(x,\ y)}{p_{\eta}(y)},\quad 当\ p_{\eta}(y)>0. \tag{21}$$

同理，在 $\xi=x$ 条件下 η 的条件分布函数为连续型，条件分布密度为

$$p_{\eta|\xi}(y|x)=\frac{p(x,\ y)}{p_{\xi}(x)},\quad 当\ p_{\xi}(x)>0. \tag{22}$$

【例 5】 设 $(\xi,\ \eta)$ 服从二元正态分布，有分布密度(10)，试求条件分布密度 $p_{\xi|\eta}(x|y)$.

由(10)及例 4，$(\xi,\ \eta)$ 的联合分布密度 $p(x,\ y)$ 及边沿分布密度 $p_{\eta}(y)$ 是连续函数，故可用公式(21)来求它的条件分

布密度.

$$
\begin{aligned}
p_{\xi|\eta}(x|y) &= \frac{p(x, y)}{p_\eta(y)} \\
&= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}\right.\right. \\
&\quad \left.\left. -\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]+\frac{(y-b)^2}{2\sigma_2^2}\right\} \\
&= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\right. \\
&\quad \left.\times\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{r^2(y-b)^2}{\sigma_2^2}\right]\right\} \\
&= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\right. \\
&\quad \left.\times\left(\frac{x-a}{\sigma_1}-r\,\frac{y-b}{\sigma_2}\right)^2\right\} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2(1-r^2)}} \exp\left\{-\frac{1}{2\sigma_1^2(1-r^2)}\right. \\
&\quad \left.\times\left[x-\left(a+\frac{\sigma_1}{\sigma_2}\,r(y-b)\right)\right]^2\right\}.
\end{aligned}
$$

从所得结果看出,在 $\eta=y$ 的条件下 ξ 的条件分布是

$$N\left(a+\frac{\sigma_1}{\sigma_2}\,r(y-b),\ \sigma_1^2(1-r^2)\right),$$

仍是正态分布. 同样可以求出 $p_{\eta|\xi}(y|x)$, 亦有类似结论. 所以,二元正态分布的条件分布仍为正态分布.

习 题 四

1. 设(ξ, η)的分布函数为 $F(x, y)$. 试用分布函数表出下列概率:

$$P(a\leqslant\xi<b,\ \eta<y),\quad P(\xi=a,\ \eta<y),$$

$$P(\xi\geqslant a,\ \eta\geqslant b),\qquad P((\xi, \eta)\in D).$$

其中 D 表示图 2-15 中给出的平面区域(阴影部分).

2. 设

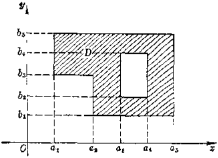


图 2-15

$$F(x, y)=\begin{cases}1, & x+y>0;\\ 0, & x+y\leqslant 0.\end{cases}$$

证明 $F(x, y)$ 满足分布函数的基本性质 1)～3)，但不满足 4)，因而不能作为分布函数.

3. 设一批产品中，A 类占 50%，B 类占 30%，C 类占 20%，从中任取 3 件，记其中 A 类的件数为 ξ_1，B 类的件数为 ξ_2，C 类的件数为 ξ_3，求 ξ_1, ξ_2, ξ_3 的联合分布.
4. 证明多项分布的边沿分布仍是多项分布.
5. 设(ξ, η)的概率分布为

η \ ξ	0	1	2	3	4	5
0	0	0.01	0.03	0.05	0.07	0.09
1	0.01	0.02	0.04	0.05	0.06	0.08
2	0.01	0.03	0.05	0.05	0.05	0.06
3	0.01	0.02	0.04	0.06	0.06	0.05

求(1)ξ, η 的边沿分布；(2)在 $\xi=3$ 的条件下，η 的条件分布列；(3)在 $\eta=1$ 的条件下，ξ 的条件分布列.

6. 若(ξ, η)的分布密度为

$$p(x, y)=\begin{cases} Ae^{-(2x+y)}, & x>0,\ y>0; \\ 0, & 其它. \end{cases}$$

(1)确定常数 A; (2)求出(ξ, η)的分布函数; (3)求出ξ的边沿分布和η的边沿分布; (4)求出条件分布密度$f_{\xi|\eta}(x|y)$; (5)$P(\xi<2|\eta<1)=$?

7. 已知(ξ, η)有分布密度

$$f(x, y)=\begin{cases} x^2+\dfrac{xy}{3}, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 2; \\ 0, & 其它. \end{cases}$$

试求下列概率

$P(\xi>0.5)$, $P((\eta-1)^2<0.25)$, $P(\xi+\eta<1)$, $P(\xi<\eta)$.

第五节 随机变量的独立性

定义 设$\xi_1, \cdots, \xi_n$是n个随机变量,如果对于任意实数$x_1, \cdots, x_n$来说,

$$P(\xi_1<x_1, \cdots, \xi_n<x_n)=P(\xi_1<x_1)\cdots P(\xi_n<x_n), \quad (1)$$

则称$\xi_1, \cdots, \xi_n$是相互独立的.

设ξ_i的分布函数为$F_i(x)$,它们的联合分布函数为$F(x_1, \cdots, x_n)$,则(1)式等价于

$$F(x_1, \cdots, x_n)=F_1(x_1)\cdots F_n(x_n), \quad (2)$$

因此,通常也用"对任意实数$x_1, \cdots, x_n$来说(2)式成立"作为$\xi_1, \cdots, \xi_n$相互独立的定义.

利用第四节(16),由(1)可知,对任何$1\leqslant i_1<\cdots<i_k\leqslant n$,

$$P(\xi_{i_1}<x_{i_1}, \cdots, \xi_{i_k}<x_{i_k})=P(\xi_{i_1}<x_{i_1})\cdots P(\xi_{i_k}<x_{i_k}).$$

所以,随机变量$\xi_1, \cdots, \xi_n$相互独立,表示由它们导出的事件$\{\xi_1<x_1\}, \cdots, \{\xi_n<x_n\}$相互独立.

下面分别讨论离散型和连续型随机变量相互独立的条件.

定理1 设$\xi_1, \cdots, \xi_n$是n个离散型随机变量，则$\xi_1, \cdots, \xi_n$相互独立的充分必要条件是对任意一组可能值$x_1, \cdots, x_n$，

$$P(\xi_1=x_1, \cdots, \xi_n=x_n)=P(\xi_1=x_1)\cdots P(\xi_n=x_n). \qquad (3)$$

【证】 充分性．假定(3)成立．设ξ_i所取的一切可能值为$x_{i1}, x_{i2}, \cdots, x_{ij}, \cdots$，则对任意实数$y_1, \cdots, y_n$，

$$\begin{aligned}
&P(\xi_1<y_1, \cdots, \xi_n<y_n)\\
&=\sum_{x_{1j_1}<y_1}\cdots\sum_{x_{nj_n}<y_n}P(\xi_1=x_{1j_1}, \cdots, \xi_n=x_{nj_n})\\
&=\sum_{x_{1j_1}<y_1}\cdots\sum_{x_{nj_n}<y_n}P(\xi_1=x_{1j_1})\cdots P(\xi_n=x_{nj_n})\\
&=\sum_{x_{1j_1}<y_1}P(\xi_1=x_{1j_1})\cdots\sum_{x_{nj_n}<y_n}P(\xi_n=x_{nj_n})\\
&=P(\xi_1<y_1)\cdots P(\xi_n<y_n),
\end{aligned}$$

由定义即知$\xi_1, \cdots, \xi_n$相互独立．

必要性．假定$\xi_1, \cdots, \xi_n$相互独立，对于任意给定的一组可能值$x_1, \cdots, x_n$及任意实数$y_1, \cdots, y_n$，由(1)可得

$$\begin{aligned}
&P\left(x_1\leqslant\xi_1<x_1+\frac{1}{m}, \ \xi_2<y_2, \ \cdots, \ \xi_n<y_n\right)\\
&=P\left(\xi_1<x_1+\frac{1}{m}, \ \xi_2<y_2, \ \cdots, \ \xi_n<y_n\right)\\
&\quad-P(\xi_1<x_1, \ \xi_2<y_2, \ \cdots, \ \xi_n<y_n)\\
&=P\left(\xi_1<x_1+\frac{1}{m}\right)P(\xi_2<y_2)\cdots P(\xi_n<y_n)\\
&\quad-P(\xi_1<x_1)P(\xi_2<y_2)\cdots P(\xi_n<y_n)\\
&=P\left(x_1\leqslant\xi_1<x_1+\frac{1}{m}\right)P(\xi_2<y_2)\cdots P(\xi_n<y_n),
\end{aligned}$$

令$m\to\infty$，利用概率的上连续性，

$$\begin{aligned}
&P(\xi_1=x_1, \ \xi_2<y_2, \ \cdots, \ \xi_n<y_n)\\
&=P(\xi_1=x_1)P(\xi_2<y_2)\cdots P(\xi_n<y_n). \qquad (4)
\end{aligned}$$

利用类似方法，由(4)又可得

$$P(\xi_1=x_1,\ \xi_2=x_2,\ \xi_3<y_3,\ \cdots,\ \xi_n<y_n)$$
$$=P(\xi_1=x_1)P(\xi_2=x_2)P(\xi_3<y_3)\cdots P(\xi_n<y_n).$$

依此类推，即知(3)必成立. ▌

第四节中，例1及例3讨论的不放回抽样问题中的ξ_1, ξ_2不是相互独立的随机变量，因为

$$P(\xi_1=0,\ \xi_2=0)=\frac{6}{20},\ P(\xi_1=0)=P(\xi_2=0)=\frac{3}{5},$$

不满足(4).

第四节中，例2及例3讨论的放回抽样问题中的ξ_1, ξ_2是相互独立的随机变量，因为可以直接验证满足条件(4)(参看表2-3).

关于连续型随机变量，有

定理2 (i) 若ξ_1, $\cdots$, ξ_n分别具有分布密度$p_1(x)$, $\cdots$, $p_n(x)$且相互独立，则(ξ_1, $\cdots$, ξ_n)有联合分布密度$p(x_1,\ \cdots,\ x_n)=p_1(x_1)\cdots p_n(x_n)$.

(ii) 若ξ_1, $\cdots$, ξ_n的联合分布密度可以表成n个单变量函数的乘积

$$p(x_1,\ \cdots,\ x_n)=f_1(x_1)\cdots f_n(x_n),$$

则ξ_1, $\cdots$, ξ_n相互独立，且存在常数c_i使$c_if_i(x)$是ξ_i的分布密度.

【证】 (i) 在给定的条件下，联合分布函数

$$F(x_1,\ \cdots,\ x_n)=F_1(x_1)\cdots F_n(x_n)$$
$$=\int_{-\infty}^{x_1}p_1(u_1)du_1\cdots\int_{-\infty}^{x_n}p_n(u_n)du_n$$
$$=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p_1(u_1)\cdots p_n(u_n)du_1\cdots du_n,$$

故知(ξ_1, $\cdots$, ξ_n)有分布密度$p(x_1,\ \cdots,\ x_n)=p_1(x_1)\cdots p_n(x_n)$.

(ii) 在给定条件下，ξ_i 的分布密度为

$$p_i(x_i)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\ \cdots,\ x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_1(x_1)\cdots f_n(x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n$$

$$=f_i(x_i)\cdot\int_{-\infty}^{\infty}f_1(x_1)dx_1\cdots\int_{-\infty}^{\infty}f_{i-1}(x_{i-1})dx_{i-1}$$

$$\times\int_{-\infty}^{\infty}f_{i+1}(x_{i+1})dx_{i+1}\cdots\int_{-\infty}^{\infty}f_n(x_n)dx_n$$

$$=c_if_i(x_i).$$

注意到

$$1=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\ \cdots,\ x_n)dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}f_1(x_1)dx_1\cdots\int_{-\infty}^{\infty}f_n(x_n)dx_n,$$

可知上面的

$$c_i=\frac{1}{\int_{-\infty}^{\infty}f_i(x_i)dx_i},\quad c_1\cdots c_n=1.$$

由此知

$$p(x_1,\ \cdots,\ x_n)=c_1f_1(x_1)\cdots c_nf_n(x_n)=p_1(x_1)\cdots p_n(x),$$

从而

$$F(x_1,\ \cdots,\ x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(u_1,\ \cdots,\ u_n)du_1\cdots du_n$$

$$=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p_1(u_1)\cdots p_n(u_n)du_1\cdots du_n$$

$$=\int_{-\infty}^{x_1}p_1(u_1)du_1\cdots\int_{-\infty}^{x_n}p_n(u_n)du_n$$

$$=F_1(x_1)\cdots F_n(x_r),$$

即 $\xi_1,\ \cdots,\ \xi_n$ 相互独立. ▌

【例 1】 设 (ξ, η) 服从二元正态分布，有分布密度 $p(x, y)$ 如前节(10)所定义，则当 $r=0$ 时，ξ、η 相互独立. 因为这时

$$p(x, y)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{(x-a)^2}{2\sigma_1^2}}\cdot e^{-\frac{(y-b)^2}{2\sigma_2^2}},$$

利用定理 2(ii)即知有上述结论.

【例 2】 已知 $\xi_1, \cdots, \xi_n$ 相互独立，且 ξ_i 服从正态分布 $N(a_i, \sigma_i^2)$, $i=1, \cdots, n$, 利用定理 2(i) 即知它们有联合分布密度

$$\begin{aligned}p(x_1, \cdots, x_n)&=p_1(x_1)\cdots p_n(x_n)\\&=\frac{1}{(2\pi)^{n/2}\sigma_1\cdots\sigma_n}\exp\left\{-\frac{1}{2}\left[\frac{(x_1-a_1)^2}{\sigma_1^2}\right.\right.\\&\qquad\left.\left.+\cdots+\frac{(x_n-a_n)^2}{\sigma_2^2}\right]\right\},\end{aligned}$$

即 $(\xi_1, \cdots, \xi_n)$ 服从正态分布 $N(a, B)$：

$$a=(a_1, \cdots, a_n)', \quad B=\begin{pmatrix}\sigma_1^2 & & 0\\ & \ddots & \\ 0 & & \sigma_n^2\end{pmatrix}.$$

下面我们把独立随机变量的概念再做一点推广.

定义 称随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是独立的，如果对任何自然数 m, $\xi_1, \xi_2, \cdots, \xi_m$ 独立.

定义 称随机向量 $(\xi_{11}, \cdots, \xi_{1n_1})$, $(\xi_{21}, \cdots, \xi_{2n_2})$, $\cdots$, $(\xi_{k1}, \cdots, \xi_{kn_k})$ 独立，如果对任何实数 $x_{11}, \cdots, x_{1n_1}, x_{21}, \cdots, x_{2n_2}, \cdots, x_{k1}, \cdots, x_{kn_k}$ 有

$$\begin{aligned}&P\left(\bigcap_{i=1}^{k}\{\xi_{i1}<x_{i1}, \cdots, \xi_{in_i}<x_{in_i}\}\right)\\&=\prod_{i=1}^{k}P(\xi_{i1}<x_{i1}, \cdots, \xi_{in_i}<x_{in_i}).\end{aligned}$$

对于随机向量，也有类似于定理 1，定理 2 的定理；另外，直观

上看，若 $(\xi_{11}, \cdots, \xi_{1n_1})$，$(\xi_{21}, \cdots, \xi_{2n_2})$，…，$(\xi_{k1}, \cdots, \xi_{kn_k})$ 独立，它们的函数 $g_1(\xi_{11}, \cdots, \xi_{1n_1})$，$g_2(\xi_{21}, \cdots, \xi_{2n_2})$，…，$g_k(\xi_{k1}, \cdots, \xi_{kn_k})$ 这样一些随机变量也应当是独立的，这一点用进一步的数学工具可以证明，这里不再多说了。

习 题 五

1. (1) 设 (ξ, η) 的分布密度为

$$f(x, y)=\begin{cases}4xy, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1;\\ 0, & \text{其它},\end{cases}$$

问 ξ, η 是否相互独立？

(2) 设 (ξ, η) 的分布密度为

$$f(x, y)=\begin{cases}8xy, & 0\leqslant x\leqslant y\leqslant 1;\\ 0, & \text{其它},\end{cases}$$

问 ξ, η 是否相互独立？

2. (1) 若 (ξ, η) 在以原点为中心边长为 2 的正方形内均匀分布，问 ξ, η 是否相互独立？

(2) 若 (ξ, η) 在以原点为中心的单位圆内均匀分布，问 ξ, η 是否相互独立？

3. 设 ξ, η 相互独立，且分别有分布密度，

$$f_1(x)=\begin{cases}e^{-x}, & x>0,\\ 0, & x\leqslant 0;\end{cases}\qquad f_2(y)=\begin{cases}e^{-y}, & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

试求联合分布密度 $f(x, y)$ 及条件分布密度 $f_{\xi|\eta}(x|y)$ 及 $f_{\eta|\xi}(y|x)$。

4. 设 (ξ, η) 有分布密度

$$f(x, y)=\begin{cases}6xy(2-x-y), & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1;\\ 0, & \text{其它},\end{cases}$$

问 ξ, η 是否相互独立？

5. 设 A_1, A_2, …, A_n 是 n 个事件，令

$$\xi_i=\begin{cases}1, & \text{若 } A_i \text{ 发生};\\ 0, & \text{若 } A_i \text{ 不发生}.\end{cases}$$

试证明，A_1, A_2, …, A_n 相互独立的充分必要条件是 ξ_1, ξ_2, …, ξ_n

相互独立.

6. 设(ξ, η, ζ)有分布密度

$$p(x, y, z)=\begin{cases}\dfrac{1}{8\pi^3}(1-\sin x\sin y\sin z), & 0\leqslant x\leqslant 2\pi,\ 0\leqslant y\leqslant 2\pi, \\ & 0\leqslant z\leqslant 2\pi; \\ 0, & \text{其它}.\end{cases}$$

试证明ξ, η, ζ两两独立,但不相互独立.

第六节 随机变量的函数的分布

本节中讨论的问题是:当随机变量ξ的分布已知时,怎样求出它的函数$\eta=f(\xi)$这样的随机变量(如$a\xi+b$, ξ^2, $\cdots$)的分布. 这问题的更一般形式是:当随机变量$\xi_1, \cdots, \xi_n$的联合分布已知时,怎样求出它们的函数

$$\eta_1=g_1(\xi_1, \cdots, \xi_n), \quad \cdots, \quad \eta_m=g_m(\xi_1, \cdots, \xi_n)$$

这样一些随机变量的联合分布.

我们只讨论连续型的情形. 主要求出一些具体形式的函数的分布,从中可以看到解决这类问题的基本方法,同时也导出一些常用的重要的分布.

【例1】 设随机变量ξ具有分布密度$p(x)$, $\eta=a\xi+b$, 其中a, b是常数且$a\neq 0$, 试求$\eta=a\xi+b$的分布密度$g(y)$.

解: 记η的分布函数为$G(y)$.

$a>0$的情形:

$$G(y)=P(\eta<y)=P(a\xi+b<y)$$

$$=P\left(\xi<\frac{y-b}{a}\right)=\int_{-\infty}^{\frac{y-b}{a}} p(t)dt$$

$$=\int_{-\infty}^{y} \frac{1}{a} p\left(\frac{u-b}{a}\right)du \quad (\text{作变换 } u=at+b),$$

这样就得出

$$g(y)=\frac{1}{a}p\left(\frac{y-b}{a}\right).$$

$a<0$ 的情形:

$$G(y)=P(\eta<y)=P(a\xi+b<y)$$
$$=P\left(\xi>\frac{y-b}{a}\right)=\int_{\frac{y-b}{a}}^{+\infty}p(t)dt$$
$$=\int_{-\infty}^{y}\frac{1}{-a}p\left(\frac{u-b}{a}\right)du\quad(\text{作变换 } u=at+b),$$

得出

$$g(y)=\frac{1}{-a}p\left(\frac{y-b}{a}\right).$$

把两种情形的结果写成统一的形式就是

$$g(y)=\frac{1}{|a|}g\left(\frac{y-b}{a}\right).\tag{1}$$

现在看一个特例:设 ξ 服从正态分布 $N(a,\sigma^2)$,由(1)知 $\eta=\frac{\xi-a}{\sigma}$ 有分布密度

$$g(y)=\sigma\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left[\sigma\left(y+\frac{a}{\sigma}\right)-a\right]^2}{2\sigma^2}\right\}$$
$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}},$$

这就是说:若 $\xi\sim N(a,\sigma^2)$,则 $\frac{\xi-a}{\sigma}\sim N(0,1)$.

【例2】 设 ξ 具有分布密度 $p(x)$,试求 $\eta=\xi^2$ 的分布密度 $g(y)$.

解:记 η 的分布函数为 $G(y)$.当 $y>0$ 时,

$$G(y)=P(\eta<y)=P(\xi^2<y)=P(-\sqrt{y}<\xi<\sqrt{y})$$
$$=\int_{-\sqrt{y}}^{\sqrt{y}}p(t)dt=\int_0^{\sqrt{y}}p(t)dt+\int_{-\sqrt{y}}^{0}p(t)dt.$$

在上式的两个积分中,分别作变量变换 $t=\sqrt{u}$ 和 $t=-\sqrt{u}$,

得到

$$G(y)=\int_0^y \frac{1}{2\sqrt{u}}[p(\sqrt{u})+p(-\sqrt{u})]du.$$

当 $y\leqslant 0$ 时， $G(y)=0,$

这样就得到

$$g(y)=\begin{cases}\frac{1}{2\sqrt{y}}[p(\sqrt{y})+p(-\sqrt{y})], & y>0;\\ 0, & y\leqslant 0.\end{cases}$$

从上面的例子我们看到，在求 $\eta=f(\xi)$ 的分布时，关键的一步是把事件 $\{f(\xi)<y\}$ 转换为 ξ 在一定范围取值的形式，从而可以利用 ξ 的分布求出它的概率。

下面转到随机向量的函数的分布。

【例 3】(和的分布) 设 (ξ_1, ξ_2) 具有分布密度 $p(x_1, x_2)$，则 $\eta=\xi_1+\xi_2$ 具有分布密度

$$g(y)=\int_{-\infty}^{\infty} p(u, y-u)du=\int_{-\infty}^{\infty} p(y-u, u)du. \qquad (2)$$

特别，当 ξ_1，ξ_2 相互独立时，设它们的分布密度分别为 $p_1(x_1)$，$p_2(x_2)$，这时 $\eta=\xi_1+\xi_2$ 的分布密度为

$$g(y)=\int_{-\infty}^{\infty} p_1(u)p_2(y-u)du$$
$$=\int_{-\infty}^{\infty} p_1(y-u)p_2(u)du. \qquad (3)$$

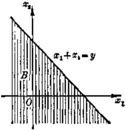


图 2-16

【证】 $\{\eta<y\}=\{\xi_1+\xi_2<y\}=\{(\xi_1, \xi_2)\in B\}$，

$$B=\{(x_1, x_2)\,|\,x_1+x_2<y\}.$$

“$\xi_1+\xi_2<y$” 这个事件表示 (ξ_1, ξ_2) 作为一个点，它的两个坐标之和小于 y，所以相当于事件“(ξ_1, ξ_2) 落在区域 B 内”(参看图 2-16)。因此，它的概率等于在此区域上的积分，

$$P(\eta<y)=\iint\limits_{x_1+x_2<y} p(x_1,\ x_2)dx_1dx_2$$
$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{y-x_1} p(x_1,\ x_2)dx_2\right]dx_1$$
$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{y} p(x_1,\ z-x_1)dz\right]dx_1$$
$$=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty} p(x_1,\ z-x_1)dx_1\right]dz,$$

由此可见 $\xi_1+\xi_2$ 有分布密度(2)．当 ξ_1, ξ_2 相互独立时，$p(x_1, x_2)=p_1(x_1)p_2(x_2)$，代入(2)即得(3)． ▌

【例 4】(商的分布) 设 (ξ_1, ξ_2) 具有分布密度 $p(x_1, x_2)$，则 $\eta=\xi_1/\xi_2$ 具有分布密度

$$g(y)=\int_{-\infty}^{\infty}|u|p(yu,\ u)du. \tag{4}$$

特别，当 ξ_1, ξ_2 相互独立时，设它们的分布密度分别为 $p_1(x_1)$, $p_2(x_2)$，则 $\eta=\xi_1/\xi_2$ 具有分布密度

$$g(y)=\int_{-\infty}^{\infty}|u|p_1(yu)p_2(u)du. \tag{5}$$

【证】 $P(\eta<y)=P(\xi_1/\xi_2<y)$

$$=\iint\limits_{x_1/x_2<y} p(x_1,\ x_2)dx_1dx_2$$
$$=\int_0^{\infty}\int_{-\infty}^{yx_2} p(x_1,\ x_2)dx_1dx_2$$
$$+\int_{-\infty}^{0}\int_{yx_2}^{\infty} p(x_1,\ x_2)dx_1dx_2$$
$$=\int_0^{\infty}\int_{-\infty}^{y} p(zx_2,\ x_2)x_2dzdx_2$$
$$-\int_{-\infty}^{0}\int_{-\infty}^{y} p(zx_2,\ x_2)x_2dzdx_2$$
$$=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty} p(zx_2,\ x_2)|x_2|dx_2\right]dz,$$

由此即知 $\eta=\xi_1/\xi_2$ 具有分布密度(4). ▌

现在，我们进一步讨论随机向量的变换，即已知$(\xi_1, \cdots, \xi_n)$有分布密度 $p(x_1, \cdots, x_n)$，求

$$\eta_1=f_1(\xi_1, \cdots, \xi_n), \quad \cdots, \quad \eta_m=f_m(\xi_1, \cdots, \xi_n)$$

的(联合)分布密度.

事件 $\{\eta_1<y_1, \cdots, \eta_m<y_m\}=\{f_1(\xi_1, \cdots, \xi_n)<y_1, \cdots, f_m(\xi_1, \cdots, \xi_n)<y_m\}$等价于："$(\xi_1, \cdots, \xi_n)$落在区域 D 内"，

$$D=\{(x_1, \cdots, x_n)\,|\,f_1(x_1, \cdots, x_n)<y_1, \cdots, f_n(x_1, \cdots, x_n)<y_m\}.$$

所以，$(\eta_1, \cdots, \eta_m\}$的分布函数

$$\begin{aligned}G(y_1, \cdots, \eta_m)&=P(\eta_1<y_1, \cdots, \eta_m<y_m)\\&=\int\cdots\int_D p(x_1, \cdots, x_n)dx_1\cdots dx_n\\&=\int\cdots\int_{\substack{f_1(x_1,\cdots,x_n)<y_1\\ \cdots\cdots\cdots\cdots\\ f_m(x_1,\cdots,x_n)<y_m}} p(x_1, \cdots, x_n)dx_1\cdots dx_n.\end{aligned}$$

接下去的问题就是要计算上式右端的积分了. 这样看来，求随机变量的函数的分布都可以按上述的步骤来进行. 当然在 p, f_i 比较复杂时会遇到计算上的困难. 在比较简单的情况下（如例 1～4），特别，当函数 $z_i=f_i(x_1, \cdots, x_n)$, $i=1, \cdots, m$, 有单值反函数(这时应有 $m=n$)，且它们的一阶偏导数存在且连续时，可以通过变量变换算出. 这方面的详细叙述可参看有关书籍(例如,王梓坤著,概率论基础及其应用,§ 2.8). 下面我们结合一些具体的例子来说明这个方法.

【例 5】 设(ξ_1, ξ_2)具有分布密度 $p(x_1, x_2)$，求

$$\eta_1=\xi_1, \quad \eta_2=\xi_1\xi_2$$

的联合分布密度 $q(y_1, y_2)$.

解：$P(\eta_1<y_1,\ \eta_2<y_2)=\iint\limits_{\substack{x_1<y_1\\ x_1x_2<y_2}} p(x_1,\ x_2)dx_1dx_2,$

作变量变换 $z_1=x_1,\ z_2=x_1x_2$. 有反函数 $x_1=z_1,\ x_2=z_2/z_1$, 而雅可比行列式

$$J=\begin{vmatrix}\dfrac{\partial x_1}{\partial z_1} & \dfrac{\partial x_1}{\partial z_2}\\ \dfrac{\partial x_2}{\partial z_1} & \dfrac{\partial x_2}{\partial z_2}\end{vmatrix}=\begin{vmatrix}1 & 0\\ -\dfrac{z_2}{z_1^2} & \dfrac{1}{z_1}\end{vmatrix}=\frac{1}{z_1},\quad |J|=\frac{1}{|z_1|},$$

代入积分得

$$P(\eta_1<y_1,\ \eta_2<y_2)=\int_{-\infty}^{y_1}\int_{-\infty}^{y_2} p\left(z_1,\ \frac{z_2}{z_1}\right)\frac{1}{|z_1|}\,dz_1dz_2,$$

于是知$(\eta_1,\ \eta_2)$的分布密度为

$$q(y_1,\ y_2)=p(y_1,\ y_2/y_1)\cdot\frac{1}{|y_1|}. \tag{6}$$

这里 $\eta_2=\xi_1\xi_2$ 是 $\xi_1,\ \xi_2$ 的乘积，它的分布可以由(6)利用求边沿分布的公式得出. 当然乘积的分布可以像例3, 例4那样来求，这里我们又提供了一个不同的方法.

乘积的分布 若 $\xi_1,\ \xi_2$ 具有分布密度 $p(x_1,\ x_2)$, 则 $\xi_1\xi_2$ 具有分布密度

$$g(y)=\int_{-\infty}^{\infty} p(u,\ y/u)\frac{1}{|u|}\,du. \tag{7}$$

【例 6】 设$(\xi_1,\ \xi_2)$相互独立，且分别服从正态分布$N(0,\ \sigma^2)$, 试求 $\eta_1=\sqrt{\xi_1^2+\xi_2^2},\ \eta_2=\dfrac{\xi_1}{\xi_2}$ 的联合分布.

记$(\xi_1,\ \xi_2)$的分布密度 $f(x_1,\ x_2)=\dfrac{1}{2\pi\sigma^2}e^{-\frac{x_1^2+x_2^2}{2\sigma^2}}$, 可以写出$(\eta_1,\ \eta_2)$的分布函数为

$$G(z_1,\ z_2)=\iint\limits_{\substack{\sqrt{x_1^2+x_2^2}<z_1\\ \frac{x_1}{x_2}<z_2}} f(x_1,\ x_2)dx_1dx_2,\quad z_1>0.$$

作变换

$$y_1=\sqrt{x_1^2+x_2^2},\quad y_2=\frac{x_1}{x_2},$$

解得两个反函数(分别对应于 $x_2>0$ 和 $x_2<0$)

$$\begin{cases} x_1=\dfrac{y_1y_2}{\sqrt{1+y_2^2}}, \\ x_2=\dfrac{y_1}{\sqrt{1+y_2^2}}; \end{cases}$$

$$\begin{cases} x_1=\dfrac{-y_1y_2}{\sqrt{1+y_2^2}}, \\ x_2=\dfrac{-y_1}{\sqrt{1+y_2^2}}. \end{cases}$$

经计算,雅可比行列式的绝对值同为

$$|J|=\frac{y_1}{1+y_2^2},\quad y_1>0.$$

变换后得到

$$\begin{aligned} G(z_1,\ z_2)=&\iint\limits_{\substack{\sqrt{x_1^2+x_2^2}<z_1\\ \frac{x_1}{x_2}<z_2\\ x_2>0}}\frac{1}{2\pi\sigma^2}e^{-\frac{x_1^2+x_2^2}{2\sigma^2}}dx_1dx_2\\ &+\iint\limits_{\substack{\sqrt{x_1^2+x_2^2}<z_1\\ \frac{x_1}{x_2}<z_2\\ x_2<0}}\frac{1}{2\pi\sigma^2}e^{-\frac{x_1^2+x_2^2}{2\sigma^2}}dx_1dx_2\\ &=\frac{1}{\pi\sigma^2}\int_0^{z_1}\int_{-\infty}^{z_2}e^{-\frac{y_1^2}{2\sigma^2}}\cdot\frac{y_1}{1+y_2^2}dy_1dy_2. \end{aligned}$$

故得$(\eta_1,\ \eta_2)$的分布密度

$$p(y_1,\ y_2)=\begin{cases} \dfrac{y_1}{\sigma^2}e^{-\frac{y_1^2}{2\sigma^2}}\cdot\dfrac{1}{\pi(1+y_2^2)}, & y_1>0, \\ 0, & y_1\leqslant 0. \end{cases}$$

从所得到的结果及第五节定理 2(ii) 可以看出 η_1 与 η_2 相互独立,并且分别具有分布密度

$$p_{\eta_1}(y_1)=\begin{cases}\dfrac{1}{\sigma^2}y_1e^{-\frac{y_1^2}{2\sigma^2}}, & y_1>0,\\ 0, & y_1\leqslant 0;\end{cases}$$

$$p_{\eta_2}(y_2)=\frac{1}{\pi(1+y_2^2)}, \quad -\infty<y_2<\infty.$$

【例 7】 设 $\xi=(\xi_1,\ \xi_2,\ \cdots,\ \xi_n)'\sim N_n(a,\ B)$, $b=(b_1,\ b_2,\ \cdots,\ b_n)'$ 是 n 维列向量, $A'=(a_{(1)},\ a_{(2)},\ \cdots,a_{(n)})$ 是 n 阶非奇异方阵, $a_{(i)}$ 是 A' 的第 i 列,试求 n 维随机向量 $A\xi+b=(a'_{(1)}\xi+b_1,\ a'_{(2)}\xi+b_2,\ \cdots,\ a'_{(n)}\xi+b_n)'$ 的联合分布。

解: 对于任意给定的实数 $z_1,\ z_2,\ \cdots,\ z_n$,

$$P(a'_{(1)}\xi+b_1<z_1,\ a'_{(2)}\xi+b_2<z_2,\ \cdots,\ a'_{(n)}\xi+b_n<z_n)$$

$$=\underset{\substack{a'_{(1)}x+b_1<z_1\\ a'_{(2)}x+b_2<z_2\\ \cdots\cdots\\ a'_{(n)}x+b_n<z_n}}{\int\cdots\int}\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}(x-a)'B^{-1}(x-a)\right\}dx_1\cdots dx_n.$$

令
$$y=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix}a'_{(1)}x+b_1\\ a'_{(2)}x+b_2\\ \vdots\\ a'_{(n)}x+b_n\end{pmatrix}=Ax+b,$$

则 $x=A^{-1}(y-b)$, $x-a=A^{-1}[y-(Aa+b)]$, 变换的雅可比行列式是

$$\left|\frac{\partial(x_1,\ x_2,\ \cdots,\ x_n)}{\partial(y_1,\ y_2,\ \cdots,\ y_n)}\right|=|A^{-1}|=|A|^{-1}.$$

所以

$$P(a'_{(1)}\xi+b_1<z_1,\ \cdots,\ a'_{(n)}\xi+b_n<z_n)$$

$$=\int_{-\infty}^{z_1}\cdots\int_{-\infty}^{z_n}\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}[y-(Aa+b)]'(A^{-1})'B^{-1}A^{-1}\right.$$

$$\left.\times[y-(Aa+b)]\right\}\cdot|A|^{-1}dy_1\cdots dy_n$$

$$=\int_{-\infty}^{z_1}\cdots\int_{-\infty}^{z_n}\frac{1}{(2\pi)^{\frac{n}{2}}|ABA'|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}[y-(Aa+b)]'(ABA')^{-1}\right.$$

$$\left.\times[y-(Aa+b)]\right\}dy_1\cdots dy_n.$$

因为 A 是非奇异方阵，ABA' 仍是正定方阵，从 $A\xi+b$ 的概率分布密度知，$A\xi+b$ 仍服从 n 维正态分布，对应的参数是 $Aa+b$ 和 ABA'.

【例 8】 (i) 设 $\xi=(\xi_1,\ \xi_2,\ \cdots,\ \xi_n)'$ 的分量是 n 个独立标准正态随机变量，B 是 n 阶正定方阵，a 是 n 维列向量. 因为 B 正定，一定有非奇异方阵 C 使 $B=CC'$. 考虑随机向量 $\eta=C\xi+a$. 由于 $\xi\sim N(0,\ I)$，由例 7 知，$\eta\sim N(a,\ CIC')$，即 $\eta\sim N(a,\ B)$.

这个结论说明：n 个相互独立的标准正态随机变量经过适当的线性变换可以得出具有任意给定参数 a, B 的正态分布 $N(a,\ B)$ 的随机向量.

(ii) 设 $\xi=(\xi_1,\cdots,\xi_n)'\sim N(a,\ B)$, $B=CC'$(C 非奇异)，则存在 n 个独立的标准正态随机变量 $\eta_1,\ \cdots,\ \eta_n$，使得 $\xi=C\eta+a$，其中 $\eta=(\eta_1,\ \cdots,\ \eta_n)'$.

事实上，只需取 $\eta=C^{-1}\xi-C^{-1}a$. 因为，由例 7，$\eta=(\eta_1,$

$\cdots, \eta_n)$服从$N(C^{-1}a-C^{-1}a, C^{-1}BC'^{-1})$，即$N(0, I)$，而且$C\eta+a=C(C^{-1}\xi-C^{-1}a)+a=\xi$.

这个结论说明：任意具有n维正态分布的随机向量总是由n个独立的标准正态随机变量经过一定的线性变换而产生的.

【例 9】 独立的标准正态随机变量经正交变换后仍是独立的标准正态随机变量. 确切地说，设$\xi=(\xi_1, \cdots, \xi_n)'$，$\eta=(\eta_1, \cdots, \eta_n)'$，若$\xi_1, \cdots, \xi_n$是独立的标准正态随机变量，$C$是一$n\times n$正交阵，$\eta=C\xi$，则$\eta_1, \cdots, \eta_n$也是独立的标准正态随机变量.

事实上，$\xi\sim N(0, I)$，根据例 7，$\eta\sim N(0, CIC')$注意到C是正交阵，即知$\eta\sim N(0, I)$.

【例 10】 $\xi=(\xi_1, \cdots, \xi_p, \xi_{p+1}, \cdots, \xi_n)'\sim N_n(a, B)$，试求$\xi_{(1)}=(\xi_1, \cdots, \xi_p)'$的分布.

解：将正定矩阵B写成分块矩阵的形式

$$B=\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

B_{11}是p阶方阵，由于B正定，B_{11}仍正定. 相应地把ξ, a, x也写成分块形式

$$\xi=\begin{pmatrix} \xi_{(1)} \\ \xi_{(2)} \end{pmatrix},\quad a=\begin{pmatrix} a_{(1)} \\ a_{(2)} \end{pmatrix},\quad x=\begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix},$$

$a_{(1)}$, $x_{(1)}$是p维列向量.

考虑随机向量

$$\begin{pmatrix} \xi_{(1)} \\ \xi_{(2)}-B_{21}B_{11}^{-1}\xi_{(1)} \end{pmatrix}=\begin{pmatrix} I_p & 0 \\ -B_{21}B_{11}^{-1} & I_{n-p} \end{pmatrix}\begin{pmatrix} \xi_{(1)} \\ \xi_{(2)} \end{pmatrix}.$$

由例 7 知，它仍服从n维正态分布，对应的参数是

$$\begin{pmatrix} I_p & 0 \\ -B_{21}B_{11}^{-1} & I_{n-p} \end{pmatrix}\begin{pmatrix} a_{(1)} \\ a_{(2)} \end{pmatrix}=\begin{pmatrix} a_{(1)} \\ a_{(2)}-B_{21}B_{11}^{-1}a_{(1)} \end{pmatrix},$$

$$\begin{pmatrix} I_p & 0 \\ -B_{21}B_{11}^{-1} & I_{n-p} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} I_p & -B_{11}^{-1}B_{12} \\ 0 & I_{n-p} \end{pmatrix}$$
$$=\begin{pmatrix} B_{11} & 0 \\ 0 & B_{22}-B_{21}B_{11}^{-1}B_{12} \end{pmatrix},$$

所以有联合分布密度

$$p(x_1, x_2, \cdots, x_n)$$
$$=\frac{1}{(2\pi)^{\frac{n}{2}}|B_{11}|^{\frac{1}{2}}|B_{22}-B_{21}B_{11}^{-1}B_{12}|^{\frac{1}{2}}}$$
$$\times\exp\left\{-\frac{1}{2}\begin{pmatrix} x_{(1)}-a_{(1)} \\ x_{(2)}-a_{(2)}+B_{21}B_{11}^{-1}a_{(1)} \end{pmatrix}'\right.$$
$$\times\begin{pmatrix} B_{11} & 0 \\ 0 & B_{22}-B_{21}B_{11}^{-1}B_{12} \end{pmatrix}^{-1}$$
$$\left.\times\begin{pmatrix} x_{(1)}-a_{(1)} \\ x_{(2)}-a_{(2)}+B_{21}B_{11}^{-1}a_{(1)} \end{pmatrix}\right\}$$
$$=\frac{1}{(2\pi)^{\frac{p}{2}}|B_{11}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x_{(1)}-a_{(1)})'B_{11}^{-1}(x_{(1)}-a_{(1)})\right\}$$
$$\times\frac{1}{(2\pi)^{\frac{n-p}{2}}|B_{22}-B_{21}B_{11}^{-1}B_{12}|^{\frac{1}{2}}}$$
$$\times\exp\left\{-\frac{1}{2}(x_{(2)}-a_{(2)}+B_{21}B_{11}^{-1}a_{(1)})'\right.$$
$$\left.\times(B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1}(x_{(2)}-a_{(2)}+B_{21}B_{11}^{-1}a_{(1)})\right\},$$

$\xi_{(1)}=(\xi_1, \cdots, \xi_p)'$ 的分布密度就是

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p(x_1, \cdots, x_n)dx_{p+1}\cdots dx_n$$
$$=\frac{1}{(2\pi)^{\frac{p}{2}}|B_{11}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x_{(1)}-a_{(1)})'B_{11}^{-1}(x_{(1)}-a_{(1)})\right\}.$$

可见，$\xi_{(1)}=(\xi_1, \cdots, \xi_p)'$ 服从 p 维正态分布，对应的参数是 $a_{(1)}$——a 的第 1，第 2，…，第 p 个分量组成的向量，B_{11}——B 的第 1 行，第 2 行，…，第 p 行与第 1 列，第 2 列，…，第 p 列相交处的元素组成的方阵。

类似的可以证明 ξ 的任何 k 个分量 $(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k})$ $(1\leqslant k\leqslant n,\ 1\leqslant i_1<i_2<\cdots<i_k\leqslant n)$ 是 k 维正态随机变量，对应的参数是 $\tilde{a}$——a 的第 i_1 个，第 i_2 个，…，第 i_k 个分量按其原有的相对位置组成的 k 维列向量，$\tilde{B}$——B 的第 i_1 行，i_2 行，…，i_k 行与第 i_1 列，i_2 列，…，i_k 列相交处的元素按其原有的相对位置组成的 k 阶方阵。这个结论的证明步骤是先用一个线性变换把 ξ 的第 $i_1, i_2, \cdots, i_k$ 个分量变换成前 k 个分量，然后利用例 10 就能得到这个结果。

最后我们给出几个由正态分布导出的重要的分布。这几个分布在数理统计中经常用到。

1. χ^2 分布

以下列函数为分布密度的分布称为**自由度为 n 的 χ^2 分布**，记作 $\chi^2(n)$：

$$g(x)=\begin{cases}0, & x\leqslant 0;\\ \dfrac{1}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0.\end{cases} \tag{8}$$

若随机变量 $\xi\sim\chi^2(n)$，则称 ξ 为**自由度为 n 的 χ^2 变量**，或简称 **χ^2 变量**。

注 由下面积分定义的函数

$$\Gamma(p)=\int_0^\infty x^{p-1}e^{-x}dx, \quad (p>0)$$

称为 Gamma 函数。利用分部积分容易证明

$$\Gamma(p)=(p-1)\Gamma(p-1).$$

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi},\ \Gamma(1)=1,$$

$$\Gamma(p)=(p-1)!\quad(\text{当}\ p\ \text{是正整数}).$$

容易看出，当 $n=2$ 时，(8)就是参数 $\lambda=\frac{1}{2}$ 的指数分布.（参看第三节，3.2.）

下面来说明产生 χ^2 分布的一个重要背景.

设 $\xi_1, \cdots, \xi_n$ 是相互独立随机变量，$\xi_i \sim N(0, 1)$，则 $\chi^2=\xi_1^2+\cdots+\xi_n^2 \sim \chi^2(n)$.

事实上，χ^2 的分布函数

$$G(x)=P(\chi^2<x)=\int\limits_{x_1^2+\cdots+x_n^2<x}\cdots\int \frac{1}{(2\pi)^{n/2}}e^{-\frac{x_1^2+\cdots+x_n^2}{2}}dx_1\cdots dx_n.$$

为了求出 χ^2 的分布密度，进行球坐标变换

$$x_1=\rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-1},$$

$$x_2=\rho\cos\theta_1\cos\theta_2\cdots\sin\theta_{n-1},$$

$$\cdots\cdots$$

$$x_{n-1}=\rho\cos\theta_1\sin\theta_2,$$

$$x_n=\rho\sin\theta_1,$$

$$J=\begin{vmatrix}\frac{\partial x_1}{\partial\rho} & \frac{\partial x_1}{\partial\theta_1} & \cdots & \frac{\partial x_1}{\partial\theta_{n-1}}\\ \frac{\partial x_2}{\partial\rho} & \frac{\partial x_2}{\partial\theta_1} & \cdots & \frac{\partial x_2}{\partial\theta_{n-1}}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial x_n}{\partial\rho} & \frac{\partial x_n}{\partial\theta_1} & \cdots & \frac{\partial x_n}{\partial\theta_{n-1}}\end{vmatrix}$$

$$=\rho^{n-1}\cdot D(\theta_1, \cdots, \theta_{n-1}),$$

可得

$$G(x)=\int_{-\pi}^{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{0}^{\sqrt{x}}\frac{1}{(2\pi)^{n/2}}e^{-\frac{\rho^2}{2}}\rho^{n-1}$$

$$\times|D(\theta_1,\ \cdots,\ \theta_{n-1})|d\rho\, d\theta_1\cdots d\theta_{n-1}$$

$$=C\int_0^x t^{\frac{n}{2}-1}e^{-\frac{t}{2}}dt,\quad (t=\rho^2)$$

其中

$$C=\int_{-\pi}^{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{2(2\pi)^{n/2}}$$

$$\times|D(\theta_1,\ \cdots,\ \theta_{n-1})|d\theta_1\cdots d\theta_{n-1}$$

是一常数．为了确定 C 的值，

$$1=\lim_{x\to+\infty}G(x)=C\int_0^{\infty}t^{\frac{n}{2}-1}e^{-\frac{t}{2}}dt$$

$$=C2^{\frac{n}{2}}\int_0^{\infty}u^{\frac{n}{2}-1}e^{-u}du=C2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right),$$

故知

$$C=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)},$$

$$G(x)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^x u^{\frac{n}{2}-1}e^{-\frac{u}{2}}du,\quad 当\ x>0.$$

2．t 分布

设 $\xi\sim N(0,\ 1)$，$\eta\sim\chi^2(n)$，且 $\xi,\ \eta$ 相互独立，则 $\zeta=\xi/\sqrt{\frac{\eta}{n}}=\sqrt{n}\,\xi/\sqrt{\eta}$ 具有分布密度

$$g(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(\frac{x^2}{n}+1\right)^{-\frac{n+1}{2}}. \tag{9}$$

这个结论可推导如下：

ξ 和 η 分别有分布密度 $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ $(-\infty<x<\infty)$ 及

$\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}}(y>0)$，且它们相互独立，所以$\zeta$的分布函数为

$$G(z)=P(\zeta<z)=P(\sqrt{n}\xi/\sqrt{\eta}<z)$$

$$=\iint\limits_{\sqrt{n}x/\sqrt{y}<z}\frac{1}{\sqrt{2\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}(x^2+y)}dx\,dy.$$

作变量变换 $\begin{cases}u=\sqrt{n}x/\sqrt{y},\\ v=y,\end{cases}$ $\begin{cases}x=n^{-\frac{1}{2}}uv^{\frac{1}{2}},\\ y=v,\end{cases}$

$$\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}n^{-\frac{1}{2}}v^{\frac{1}{2}} & \frac{1}{2}n^{-\frac{1}{2}}uv^{-\frac{1}{2}}\\ 0 & 1\end{vmatrix}=n^{-\frac{1}{2}}v^{\frac{1}{2}},$$

得 $$G(z)=\int_{-\infty}^{z}\int_{0}^{\infty}\frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}v^{\frac{n-1}{2}}e^{-\frac{1}{2}\left(\frac{u^2}{n}+1\right)v}dv\,du,$$

由此知ζ有分布密度

$$g(z)=\frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_{0}^{\infty}v^{\frac{n-1}{2}}e^{\frac{-1}{2}\left(\frac{z^2}{n}+1\right)v}dv.$$

再作变换$t=\frac{1}{2}\left(\frac{z^2}{n}+1\right)v$，

$$g(z)=\frac{1}{\sqrt{2n\pi}2^{n/2}\Gamma\left(\frac{n}{2}\right)}$$

$$\times\int_{0}^{\infty}2^{\frac{n+1}{2}}\left(\frac{z^2}{n}+1\right)^{-\frac{n+1}{2}}t^{\frac{n+1}{2}-1}e^{-t}dt$$

$$=\frac{1}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(\frac{z^2}{n}+1\right)^{-\frac{n+1}{2}}\cdot\Gamma\left(\frac{n+1}{2}\right).$$

这就是(9).

$\zeta=\xi\Big/\sqrt{\frac{\eta}{n}}$ 的分布密度也可以按下述方法求出：先求出 $\sqrt{\frac{\eta}{n}}$ 的分布密度，再利用商的分布（例 4）求出 ζ 的分布密度.

以(9)为分布密度的分布称为自由度为 n 的 t 分布，简记作 $t(n)$. 若随机变量 $\xi\sim t(n)$ 则称 ξ 是自由度为 n 的 t 变量，或简称 t 变量.

3. F 分布

设 $\xi\sim\chi^2(m)$, $\eta\sim\chi^2(n)$, 且 ξ, η 相互独立，则 $\zeta=\frac{\xi}{m}\Big/\frac{\eta}{n}=\frac{n}{m}\cdot\frac{\xi}{\eta}$ 具有分布密度

$$g(x)=\begin{cases}\dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}x^{\frac{m}{2}-1}\,(mx+n)^{-\frac{m+n}{2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}\tag{10}$$

我们通过下述步骤来导出 ζ 的分布密度：先求出 ξ/η 的分布密度，再进一步求出 $\zeta=\frac{n}{m}\,\frac{\xi}{\eta}$ 的分布密度.

ξ, η 相互独立，且分别具有分布密度

$$p_1(x)=\frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}x^{\frac{m}{2}-1}e^{-\frac{x}{2}},\quad x>0;$$

$$p_2(y)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}},\quad y>0.$$

利用商的分布密度公式(5)（例 4），可得 ξ/η 的分布密度

$$p_{\xi/\eta}(z)=\int_{-\infty}^{\infty} p_1(zu)\,p_2(u)\,|u|\,du$$

$$=\int_0^{\infty}\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}z^{\frac{m}{2}-1}u^{\frac{m+n}{2}-1}e^{-\frac{zu+u}{2}}\,du.$$

作变换 $t=\frac{1}{2}(z+1)u$,

$$p_{\xi/\eta}(z)=\int_0^{\infty}\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$\times z^{\frac{m}{2}-1}2^{\frac{m+n}{2}}(z+1)^{-\frac{m+n}{2}}t^{\frac{m+n}{2}-1}e^{-t}dt$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}z^{\frac{m}{2}-1}(z+1)^{-\frac{m+n}{2}},\quad z>0.$$

由此利用例 1 中(1), 即知 $\zeta=\frac{n}{m}\frac{\xi}{\eta}$ 的分布密度为

$$g(z)=\frac{m}{n}\,p_{\xi/\eta}\left(\frac{m}{n}z\right)$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{m}{n}\left(\frac{m}{n}z\right)^{\frac{m}{2}-1}\left(\frac{m}{n}z+1\right)^{-\frac{m+n}{2}}$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}z^{\frac{m}{2}-1}(mz+n)^{-\frac{m+n}{2}},$$

$$z>0.$$

以(10)为分布密度的分布称为第一自由度为 m, 第二自由度为 n 的 F 分布, 简记作 $F(m,\ n)$. 若随机变量 $\xi\sim F(m,$

n)，则称 ξ 是自由度为 m、n 的 F 变量，或简称 F 变量.

χ^2 分布，t 分布和 F 分布是数理统计中经常用到的三个分布，它们的数值已制成表以供查用.

习 题 六

1．设 ξ 的分布列为

ξ	-2	-1	0	1	2
p	0.15	0.2	0.3	0.2	0.15

试分别求出 $\eta=\xi^2$, $\zeta=2\xi-1$, $\psi=|\xi|+1$ 的分布列.

2．设 ξ, η 的分布列分别为

ξ	0	1	2
p	0.5	0.3	0.2

η	0	2
p	0.6	0.4

且 ξ, η 相互独立．试求 $\zeta=\xi+\eta$ 的分布列和 $\psi=\xi\eta$ 的分布列.

3．若 ξ_1, ξ_2 相互独立，分别服从参数为 λ_1, λ_2 的泊松分布，证明

(1) $\xi_1+\xi_2$ 具有参数为 $\lambda_1+\lambda_2$ 的泊松分布；

(2) $P(\xi_1=k|\xi_1+\xi_2=n)=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$.

4．设 ξ_1, ξ_2 相互独立，ξ_1 服从二项分布 $b(n_1, p)$，ξ_2 服从二项分布 $b(n_2, p)$．试证，$\xi_1+\xi_2$ 服从二项分布 $b(n_1+n_2, p)$.

设在 n 次独立试验中(每次成功的概率为 p)，令 μ_i 是第 i 次试验中成功的次数，利用上述结论证明 $\mu_1+\mu_2+\cdots+\mu_n$ 服从二项分布 $b(n, p)$.

5．设 $\xi\sim N(a, \sigma^2)$，证明 $\eta=e^\xi$ 具有分布密度

$$p(x)=\frac{1}{x\sqrt{2\pi}\sigma}e^{-\frac{(\ln x-a)^2}{2\sigma^2}},\quad 0<x<\infty.$$

这个分布称为**对数正态分布**.

6．设 ξ 服从指数分布，具有分布密度

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0;\\ 0, & x<0.\end{cases}$$

证明，$\eta=\xi^{\frac{1}{a}}$（$a>0$ 是一常数）具有分布密度

$$p_\eta(x)=\begin{cases}a\lambda x^{a-1}e^{-\lambda x^a}, & x\geqslant 0;\\ 0, & x<0.\end{cases}$$

这个分布称为韦布（Weibull）分布。

7．设对球的直径作近似测量，测量值在$[a, b]$内均匀分布，求球积体的分布密度。

8．若气体分子的速度是随机向量 $V=(\xi_1, \xi_2, \xi_3)$，各分量相互独立，且均服从正态分布 $N(0, \sigma^2)$。试证速度的绝对值 $\xi=\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}$ 服从马克斯威分布

$$f(x)=\sqrt{\frac{2}{\pi}}\frac{x^2}{\sigma^3}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x>0.$$

又分子能量 $\eta=\frac{1}{2}m\xi^2$ 的分布密度为何？

9．已知方程 $x^2+\alpha x+\beta=0$ 的两根独立地分别在 $[-1, 1]$ 中均匀分布，试求系数 α, β 的分布密度。

10．设(ξ, η, ζ)的分布密变为

$$p(x, y, z)=\begin{cases}6(1+x+y+z)^{-4}, & x>0,\ y>0,\ z>0;\\ 0, & \text{其它}.\end{cases}$$

求 $\xi+\eta+\zeta$ 的分布密度。

11．若 ξ, η 独立，$\xi_1\sim\chi^2(m)$，$\xi_2\sim\chi^2(n)$，证明 $\xi+\eta\sim\chi^2(m+n)$。

12．已知 ξ, η 相互独立，均具有分布密度

$$p(x)=\begin{cases}e^{-x}, & x>0;\\ 0, & \text{其它}.\end{cases}$$

问 $\xi+\eta$，$\frac{\xi}{\xi+\eta}$ 是否相互独立，

13．(ξ, η)具有分布密度

$$p(x, y)=\begin{cases}\frac{1+xy}{4}, & |x|<1,\ |y|<1;\\ 0, & \text{其它}.\end{cases}$$

试证，ξ 与 η 不独立，但 ξ^2 与 η^2 独立。

14．ξ, η 独立，各服从几何分布

$$P(\xi=k)=P(\eta=k)=pq^{k-1}, \quad k=1, 2, \cdots.$$

令 $\zeta=\max(\xi,\eta)$，试求 (1)(ζ,ξ)的联合分布，(2)ζ 的分布，(3)ξ 关于 ζ 的条件分布。

[提示：$\{\zeta=k,\ \xi=l\}=\emptyset$， 当 $k<l$；

$\{\zeta=k,\ \xi=l\}=\{\eta=k,\ \xi=l\}$， 当 $k>l$；

$\{\zeta=k,\ \xi=l\}=\{\eta=1,\ \xi=l\}+\cdots+\{\eta=k,\ \xi=l\}$， 当 $k=l$；

$\{\zeta=k\}=\{\max(\xi,\ \eta)=k\}$

$=\{\xi=k,\ \eta=1\}+\cdots+\{\xi=k,\ \eta=k\}$

$+\{\xi=1,\ \eta=k\}+\cdots+\{\xi=k-1,\ \eta=k\}$.]

15. 设 $\xi_1\sim F(m,\ n)$，$\xi_2\sim F(n,\ m)$，$F_1(x)$，$F_2(x)$ 分别是 ξ_1，ξ_2 的分布函数，证明：对任何正实数 x，有

$$F_1(x)=1-F_2(1/x)$$

(这个性质有时对查 F 分布表有用)。

补 充 题

1. ξ 的分布函数 $F(x)$连续，证明 $F(\xi)$在$[0,\ 1]$上均匀分布。
2. ξ 在$[0,\ 1]$上均匀分布，证明 $-\ln\xi$ 服从自由度为 2 的 χ^2 分布。
3. ξ 服从对数正态分布，证明 ξ^r 仍服从对数正态分布。
4. $(\xi,\ \eta)$是二维随机变量，若 $P(\xi=C)=1$，C 是常数，证明 ξ，η 独立。
5. 多维正态分布的边沿分布仍是正态分布，反过来对吗？ 考察下面的例子：$(\xi,\ \eta)$有概率分布密度

$$P(x,\ y)=\begin{cases}0, & \text{当 } xy<0;\\ \dfrac{1}{\pi}e^{-\frac{1}{2}(x^2+y^2)}, & \text{当 } xy\geqslant 0.\end{cases}$$

算出 ξ，η 的边沿分布。

6. 在线段$(0,\ a)$上任意投掷两点，试求两点距离平方的分布。
7. $(\xi,\ \eta)$服从参数为$(0,\ 0;1,\ 1;\rho)$的二维正态分布，证明

$$P(\xi>0,\ \eta>0)=P(\xi<0,\ \eta<0)=\frac{1}{4}+\frac{1}{2\pi}\arcsin\rho,$$

$$P(\xi<0,\ \eta>0)=P(\xi>0,\ \eta<0)=\frac{1}{4}-\frac{1}{2\pi}\arcsin\rho.$$

8. ξ_1，ξ_2，…，ξ_n 独立，都在 $[a,\ b]$上均匀分布。设 $\eta_1=\min\limits_{1\leqslant i\leqslant n}\{\xi_i\}$，

$\eta_n = \max\limits_{1\leq i\leq n}\{\xi_i\}$，试分别求 η_1, η_n 的分布函数及分布密度.

9. 上题，试求 (η_1, η_n) 的联合分布密度.

10. 上题，试求 $\eta_1+\eta_n$ 的分布密度.

第二章小结

当随机现象用随机变量来表示时，用随机变量的分布律就可以描述随机现象的概率规律，这比一般用概率空间描述随机现象有许多方便之处，例如，我们可以用所熟悉的微积分知识来研究分布律.

研究随机变量的分布及其数字特征(见第三章)是本书的中心内容.

离散型分布和连续型分布是两类常见的分布，前者有分布列，后者有分布密度.

对独立随机变量的研究在概率论中占有重要地位. 本章着重讨论了独立随机变量函数的分布.

求随机变量的分布的问题仍是求事件概率的问题. 求随机变量的分布(包括一维随机变量的分布，多维随机变量的联合分布与边沿分布，随机变量函数的分布)是一项基本功，要认真地做一些习题才会有所收益.

正态分布，二项分布，均匀分布，泊松分布都是常用的分布，除正态分布外，它们产生的背景都在第一章讲过，要记住这些分布的形式.

χ^2 分布，t 分布，F 分布是数理统计中常用的分布，分布的形式不必记住，但要熟悉是哪些独立正态随机变量的函数服从这些分布. 为了这个目的，建议读者在学了这一部分后接着读第五章 1.3.

第三章

数字特征与特征函数

随机变量(或随机变量的分布)的数字特征是指刻划随机变量(或它的分布)某些特征的数值．打一个比喻，譬如，你想研究一个工厂的工人年龄状况，如果你能得到一张这个工厂的工人年龄登记表，你就会感到很满意，因为你需要的资料全有了．但是，如果你想比较一下两个工厂工人的年龄，例如，总的看哪一个工厂工人的年龄更大一些，那你只凭两张登记表还不行，还要进一步算出每个工厂工人的平均年龄，看一看哪一个工厂工人的平均年龄大．对年龄登记表来说，年龄平均值就是一个数字特征，它在一定意义下刻划了一个工厂工人年龄的大小；最大年龄、最小年龄、最大年龄与最小年龄的差等等也是这样的数字特征．

对于随机变量，我们在第二章中曾经说过：它的取值规律由它的分布律完全刻划了，但也需要定义一些数值用来刻划它的某些特征，如取值集中的程度、取值的大小等等．

第一节 数学期望

1.1 数学期望的定义

为了帮助读者理解数学期望的定义，先举一个例子．在一个袋子里装着六个相同的小球，其中有三个球上写着“2”，两个球上写着“1”，一个球上写着“－1”．规定从袋中任意取出一个球，记录下球上的数，然后把球放回袋中为一次试验．

如果问你在三次试验中记录下来的三个数的平均值是多少，你会觉得很难回答. 因为记录下来的三个数可能是三个“2”，也可能是两个“1”，一个“-1”等等，对这三个数是什么都很难预料，还怎么说出它们的平均值呢？这个问题的抽象提法就是：有一个分布列为

$$\begin{pmatrix} 2 & 1 & -1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

的离散型随机变量 ξ——试验的结果，因为让 ξ 独立地取三个值，将取到的第一个值 ξ_1，第二个值 ξ_2，第三个值 ξ_3 仍是随机变量，所以平均值 $\frac{1}{3}(\xi_1+\xi_2+\xi_3)$ 也是随机变量，无法精确预言它将取什么值. 然而如果让试验次数增多，即让 ξ 取比较多的值，情况就会发生变化. 例如，让 ξ 取 n_1 个值，其中有 n_{11} 个“2”，n_{12} 个“1”，n_{13} 个“-1”. 计算这 n_1 个值的平均值得：

$$2\times\frac{n_{11}}{n_1}+1\times\frac{n_{12}}{n_1}+(-1)\times\frac{n_{13}}{n_1}.$$

再让 ξ 取 n_2 个值，其中有 n_{21} 个“2”，n_{22} 个“1”，n_{23} 个“-1”. 计算这 n_2 个值的平均值得：

$$2\times\frac{n_{21}}{n_2}+1\times\frac{n_{22}}{n_2}+(-1)\times\frac{n_{23}}{n_2},$$

继续这样做下去得到 ξ 取的许多组值的平均值

$$2\times\frac{n_{31}}{n_3}+1\times\frac{n_{32}}{n_3}+(-1)\frac{n_{33}}{n_3},$$

………………………………

其中 $\frac{n_{11}}{n_1}$，$\frac{n_{21}}{n_2}$，… 分别是在 n_1，n_2，… 次试验中事件 $\{\xi=2\}$ 出现的频率. 因为 n_1，n_2，…比较大，这些频率稳定在

$$P(\xi=2)=\frac{1}{2}$$

上. 同样 $\frac{n_{12}}{n_1}$, $\frac{n_{22}}{n_2}$, … 稳定在

$$P(\xi=1)=\frac{1}{3}$$

上, $\frac{n_{13}}{n_1}$, $\frac{n_{23}}{n_2}$, … 稳定在

$$P(\xi=-1)=\frac{1}{6}$$

上,因而这些平均值稳定在

$$2\times\frac{1}{2}+1\times\frac{1}{3}+(-1)\times\frac{1}{6}=\frac{7}{6}$$

上. 情况确实发生了变化,当 ξ 取的值比较多时,平均值是稳定的. 我们将在这个意义上说 ξ 的平均值, 而数值"$\frac{7}{6}$"则表示了 ξ 的平均值的大小.

一般,若离散型随机变量 ξ 有分布列

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}$$

当 ξ 取的值比较多时,平均值将稳定在 $\sum_{i=1}^{m} a_i p_i$ 上, 我们将称数值 $\sum_{i=1}^{m} a_i p_i$ 为 ξ 的数学期望.

定义 离散型随机变量 ξ 有分布列

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}, \tag{1}$$

若级数

$$\sum_{i=1}^{\infty} |a_i| p_i \tag{2}$$

收敛，则称级数和

$$\sum_{i=1}^{\infty} a_i p_i \tag{3}$$

为随机变量 ξ 的数学期望，记作 $E\xi$，否则称 ξ 的数学期望不存在。

要求级数(3)绝对收敛，为的是使级数和与项的排列次序没有关系。这个要求是自然的，因为随机变量的平均值稳定在什么地方应当与(1)中各项排列的次序没有关系。

下面我们将要给出连续型随机变量数学期望的定义。为了帮助读者理解这个定义，仍旧先看一个具体的例子。设随机变量 ξ 只在有限区间 $[a, b)$ 内取值，有概率分布密度 $p(x)$。当 ξ 取比较多的值时，我们希望知道这些值的平均值稳定在什么地方。设 ξ 取的值为 $x_1, x_2, \cdots, x_n$，将区间 $[a, b)$ 等分为 m 个小区间，分点为 $a_1(=a), a_2, \cdots, a_{m+1}(=b)$，如果 $x_r \in [a_i, a_{i+1})$，就用 a_i 代替 x_r。假设 ξ 取的 n 个值中有 n_i 个属于 $[a_i, a_{i+1})$，$i=1, 2, \cdots, m$，则

$$\frac{1}{n}\sum_{r=1}^{n} x_r \approx \sum_{i=1}^{m} a_i \cdot \frac{n_i}{n}. \tag{4}$$

因为 $\frac{n_i}{n}$ 是事件 $\{a_i \leqslant \xi < a_{i+1}\}$ 在 n 次试验中出现的频率，稳定在 $P(a_i \leqslant \xi < a_{i+1}) \approx p(a_i)(a_{i+1}-a_i)$ 上，$i=1, 2, \cdots, m$，所以 $\sum_{i=1}^{m} a_i \frac{n_i}{n}$ 稳定在 $\sum_{i=1}^{m} a_i p(a_i)(a_{i+1}-a_i)$ 上，而这个数就是积分 $\int_a^b xp(x)dx$ 的积分和，又当 n 比较大时 m 也可以取比较大的值，(4)式两端相差甚少，所以 $\frac{1}{n}\sum_{r=1}^{n} x_r$ 应当稳定在积分 $\int_a^b xp(x)dx$ 上。

定义 连续型随机变量 ξ 有概率分布密度 $p(x)$，若积分

$$\int_{-\infty}^{+\infty}|x|p(x)dx \tag{5}$$

收敛，则称积分

$$\int_{-\infty}^{+\infty}xp(x)dx \tag{6}$$

为随机变量ξ的数学期望，记作$E\xi$，否则称ξ的数学期望不存在.

从数学期望的定义可以看到，$E\xi$是由ξ的概率分布确定的，所以也称$E\xi$为该分布的数学期望.

【例 1】 ξ服从参数为λ的泊松分布，试求$E\xi$.

解：ξ有分布列

$$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0, 1, 2, \cdots,$$

相当于(2)的级数.

$$\sum_{k=0}^{+\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}$$

是非负项级数. 因为

$$\sum_{k=0}^{+\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{+\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k-1=0}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}$$
$$=\lambda e^{-\lambda}e^{\lambda}=\lambda,$$

所以$E\xi$存在，而且$E\xi=\lambda$. 这里我们看到了泊松分布中参数λ的含义.

【例 2】 $\xi\sim N(a, \sigma^2)$，试求$E\xi$.

解：ξ有概率分布密度

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}},$$

相当于(5)的积分是

$$\int_{-\infty}^{+\infty}|x|p(x)dx.$$

由于

$$\int_{-\infty}^{+\infty}|x|p(x)dx\leqslant\int_{-\infty}^{+\infty}|a|p(x)dx+\int_{-\infty}^{+\infty}|x-a|p(x)dx$$

$$=|a|+\int_{-\infty}^{+\infty}|x-a|\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=|a|+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}|u|e^{-\frac{1}{2}u^2}du\quad\left(u=\frac{x-a}{\sigma}\right)$$

$$=|a|+\frac{2\sigma}{\sqrt{2\pi}}\int_{0}^{+\infty}ue^{-\frac{u^2}{2}}du$$

$$=|a|+\frac{2\sigma}{\sqrt{2\pi}}\int_{0}^{+\infty}e^{-\frac{u^2}{2}}d\left(\frac{u^2}{2}\right)=|a|+\frac{2\sigma}{\sqrt{2\pi}},$$

故 $E\xi$ 存在,

$$E\xi=\int_{-\infty}^{+\infty}xp(x)dx=\int_{-\infty}^{+\infty}x\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma u+a)e^{-\frac{1}{2}u^2}du\quad\left(u-\frac{x-a}{\sigma}\right)$$

$$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}ue^{-\frac{1}{2}u^2}du+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}u^2}du=a.$$

这里我们又看到了正态分布中第一个参数 a 的含义.

【例 3】 设 ξ 有概率分布密度

$$p(x)=\frac{1}{\pi(1+x^2)}.$$

这时

$$\int_{-\infty}^{+\infty}|x|\frac{1}{\pi(1+x^2)}dx=\frac{2}{\pi}\int_{0}^{+\infty}\frac{x}{1+x^2}dx$$

$$=\frac{1}{\pi}\int_{0}^{+\infty}\frac{d(1+x^2)}{1+x^2}=\frac{1}{\pi}\ln(1+x^2)\Big|_{0}^{+\infty}$$

$$=\lim_{x\to+\infty}\frac{1}{\pi}\ln(1+x^2),$$

相当于(5)的积分发散,故 $E\xi$ 不存在.

【例 4】 ξ 服从参数为(n, p)的二项分布，试求 $E\xi$.

解：ξ 有分布列

$$P(\xi=k)=C_n^k p^k(1-p)^{n-k}, \quad k=0, 1, \cdots, n.$$

因为 ξ 是只取有限个值的离散型随机变量，根据定义，$E\xi$ 一定存在.

$$\begin{aligned}
E\xi &= \sum_{k=0}^{n} kC_n^k p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= np\sum_{k-1=0}^{n-1}\frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} \\
&\quad \times p^{k-1}(1-p)^{(n-1)-(k-1)} \\
&= np[p+(1-p)]^{n-1}=np.
\end{aligned}$$

【例 5】 ξ 是自由度为 n 的 χ^2 变量，试求 $E\xi$.

解：

$$\begin{aligned}
E\xi &= \int_0^{+\infty} x\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx \\
&= \frac{2}{\Gamma\left(\frac{n}{2}\right)}\int_0^{+\infty}\left(\frac{x}{2}\right)^{\frac{n}{2}}e^{-\frac{x}{2}}d\left(\frac{x}{2}\right) \\
&= \frac{2}{\Gamma\left(\frac{n}{2}\right)}\Gamma\left(\frac{n}{2}+1\right) \\
&= \frac{2}{\Gamma\left(\frac{n}{2}\right)}\frac{n}{2}\Gamma\left(\frac{n}{2}\right)=n.
\end{aligned}$$

数学期望是概率论的一个基本概念，希望读者在学习这一部分时注意下面几点：

第一，什么叫随机变量的数学期望？由定义知，一个离散型随机变量的数学期望就是由(3)式给出的一个绝对收敛级

数的和；一个连续型随机变量的数学期望就是由(6)式给出的一个绝对可积函数的积分．

第二，谈到数学期望概念的含义时，有时说："数学期望表示随机变量的平均值"．对这句话应该理解为：当随机变量独立地取比较多的值时，这些值的平均值稳定在随机变量的数学期望上．在定义之前我们已用个别例子说明过这一点，第四章大数定律部分还将看到更一般的说明．

第三，在本书中，我们只对离散型的和连续型的随机变量给出了数学期望的概念．对一般随机变量定义数学期望，需要利用更多的数学工具，读者可参看严士健等《测度论与概率论基础》．

今后，在许多场合要讨论随机变量的函数的数学期望，这里给出一个关于随机变量函数的数学期望的重要定理．

定理 1) $(\xi_1, \xi_2, \cdots, \xi_n)$是离散型随机变量，有分布列

$$P(\xi_1=x_{1i_1}, \xi_2=x_{2i_2}, \cdots, \xi_n=x_{ni_n})=p_{i_1i_2\cdots i_n}$$
$$(i_1, i_2, \cdots, i_n=1, 2, \cdots),$$

$g(t_1, t_2, \cdots, t_n)$是一个n元函数，若级数

$$\sum_{i_1}\sum_{i_2}\cdots\sum_{i_n}|g(x_{1i_1}, x_{2i_2}, \cdots, x_{ni_n})|p_{i_1i_2\cdots i_n}$$

收敛，则$Eg(\xi_1, \xi_2, \cdots, \xi_n)$存在，并且

$$Eg(\xi_1, \xi_2, \cdots, \xi_n) =\sum_{i_1}\sum_{i_2}\cdots\sum_{i_n}g(x_{1i_1}, x_{2i_2}, \cdots, x_{ni_n})p_{i_1i_2\cdots i_n}.$$

2) $(\xi_1, \xi_2, \cdots, \xi_n)$是连续型随机变量，有概率分布密度$p(x_1, x_2, \cdots, x_n)$，$g(x_1, x_2, \cdots, x_n)$是一个$n$元函数，若积分

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}|g(x_1, \cdots, x_n)|p(x_1, \cdots, x_n)dx_1\cdots dx_n$$

收敛，则$Eg(\xi_1, \cdots, \xi_n)$存在，并且

$$Eg(\xi_1, \cdots, \xi_n)$$
$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}g(x_1, \cdots, x_n)p(x_1, \cdots, x_n)dx_1\cdots dx_n.$$

证明这个定理需要更多的数学工具，这里就不证了．需要说明的是：按照数学期望的定义，如果 $g(\xi_1, \cdots, \xi_n)$ 是离散型随机变量，应该先求出它的分布列；如果 $g(\xi_1, \cdots, \xi_n)$ 是连续型随机变量，应该先求出它的分布密度，然后判别 $Eg(\xi_1, \cdots, \xi_n)$ 是否存在，在 $Eg(\xi_1, \cdots, \xi_n)$ 存在时利用(3)或(6)计算 $Eg(\xi_1, \cdots, \xi_n)$．如果 $g(\xi_1, \cdots, \xi_n)$ 既不是离散型也不是连续型随机变量，就应该按照一般随机变量数学期望的定义判别 $Eg(\xi_1, \cdots, \xi_n)$ 是否存在及计算 $Eg(\xi_1, \cdots, \xi_n)$．这个定理的意义在于给出了一个更便于使用的判别 $Eg(\xi_1, \cdots, \xi_n)$ 是否存在及计算 $Eg(\xi_1, \cdots, \xi_n)$ 的方法．

有一些随机变量函数的数学期望有特别的名称．

定义 ξ 是随机变量，r 是非负整数，s 是非负实数，称 $E\xi^r$ 为 ξ 的 r 阶矩，称 $E(\xi-E\xi)^r$ 为 ξ 的 r 阶中心矩，称 $E|\xi|^s$ 为 ξ 的 s 阶绝对矩，称 $E|\xi-E\xi|^s$ 为 ξ 的 s 阶绝对中心矩．

多维随机变量数学期望的定义包含在下面随机矩阵数学期望的定义之中．

定义 若随机变量 $\xi_{ij}(i=1, 2, \cdots, m; j=1, 2, \cdots, n)$ 的数学期望存在，则称矩阵

$$(E\xi_{ij})=\begin{bmatrix} E\xi_{11} & E\xi_{12} & \cdots & E\xi_{1n} \\ E\xi_{21} & E\xi_{22} & \cdots & E\xi_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ E\xi_{m1} & E\xi_{m2} & \cdots & E\xi_{mn} \end{bmatrix}$$

为随机矩阵

$$(\xi_{ij})=\begin{bmatrix}\xi_{11} & \xi_{12} & \cdots & \xi_{1n}\\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \xi_{m1} & \xi_{m2} & \cdots & \xi_{mn}\end{bmatrix}$$

的数学期望，记作 $E(\xi_{ij})$.

【例 6】 $\xi\sim N(0,1)$，试求 ξ 的 k 阶矩，k 为非负整数.

解：当 k 为奇数时

$$E\xi^k=\int_{-\infty}^{+\infty}x^k\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.$$

当 k 为偶数时

$$\begin{aligned}E\xi^k&=\int_{-\infty}^{+\infty}x^k\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{2}{\sqrt{2\pi}}\int_0^{+\infty}x^ke^{-\frac{x^2}{2}}dx\\&=\frac{\sqrt{2}}{\sqrt{\pi}}\int_0^{+\infty}2^{\frac{k-1}{2}}\left(\frac{x^2}{2}\right)^{\frac{k-1}{2}}e^{-\frac{x^2}{2}}d\left(\frac{x^2}{2}\right)\\&=\frac{2^{\frac{k}{2}}}{\sqrt{\pi}}\Gamma\left(\frac{k+1}{2}\right)\\&=\frac{2^{\frac{k}{2}}}{\Gamma\left(\frac{1}{2}\right)}\frac{k-1}{2}\cdot\frac{k-3}{2}\cdots\frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\\&=(k-1)(k-3)\cdots3\cdot1.\end{aligned}$$

【例 7】 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)'$ 服从参数为 $a=(a_1,a_2,\cdots,a_n)'$，$B=(b_{ij})$（n 阶正定阵）的多维正态分布，试求 $E\xi$.

解：对任何 $i(1\leqslant i\leqslant n)$，由第二章第六节例 8 知，$\xi_i\sim N(a_i,b_{ii})$. 由例 2 知，$E\xi_i=a_i$. 所以

$$E\xi=(E\xi_1,E\xi_2,\cdots,E\xi_n)'=(a_1,a_2,\cdots,a_n)'=a.$$

【例 8】 甲、乙两台自动机床生产同一种标准件，ξ 和 η

分别表示甲，乙各生产 1000 件产品中的次品数．经过一段时间的考察，ξ、η 的分布列分别为

ξ	0	1	2	3
P_ξ	0.7	0.1	0.1	0.1

η	0	1	2	3
P_η	0.5	0.3	0.2	0.0

试问哪一台机床质量好一些？

比较两个随机变量的平均值

$$E\xi=0\times0.7+1\times0.1+2\times0.1+3\times0.1=0.6,$$

$$E\eta=0\times0.5+1\times0.3+2\times0.2+3\times0.0=0.7.$$

$E\xi<E\eta$，说明机床甲在生产的 1000 件产品中次品平均数比较小，所以机床甲质量比较好．

习 题 一

1．随机变量 ξ 有分布列

ξ	−3	0	1	5
P_3	0.1	0.2	0.3	0.4

试求 $E\xi$ 及 $E|\xi|^3$．

2．ξ 服从拉普拉斯分布，即有概率分布密度

$$p(x)=\frac{1}{2a}e^{-\frac{|x-\beta|}{a}},\quad a>0,$$

试求 $E\xi$ 及 $E\xi^2$．

3．ξ 服从马克斯威尔分布，即有概率分布密度

$$p(x)=\begin{cases}\dfrac{4x^2}{a^3\sqrt{\pi}}e^{-\frac{x^2}{a^2}}, & \text{当 } x>0;\\ 0, & \text{当 } x\leqslant 0.\end{cases}$$

试求 $E\xi$ 及 $E\xi^2$.

4. 袋中装有3个白球3个红球，将球一个一个取出，每次取出后不放回. 设在第 ξ 次第二次取得红球，试求 $E\xi$.

5. 上题中如果白球个数为 n，试求 $E\xi$.

6. 在线段 $(0, l)$ 上任意投两个点，试求其间距离 n 次方的数学期望.

7. ξ 是取非负整数值的随机变量，若数学期望存在，则

$$E\xi=\sum_{k=1}^{+\infty}P(\xi\geqslant k).$$

8. $\xi\sim N(a, \sigma^2)$，试求 ξ 的 k 阶中心矩.

1.2 数学期望的性质

性质 1 一个常数的数学期望就等于这个常数.

【证】 可以把常数 c 看成一个离散型随机变量，有分布列 $P(c=c)=1$. 由定义 1，$Ec=c$. 】

性质 2 c 是一个常数. 若随机变量 ξ 的数学期望 $E\xi$ 存在，则 $c\xi$ 的数学期望存在，并且 $E(c\xi)=cE\xi$.

【证】 在本书中只能分别对离散型和连续型的情形给出证明.

若 ξ 是离散型随机变量，有分布列

$$P(\xi=a_i)=p_i,\quad i=1, 2, \cdots.$$

这时级数

$$\sum_{i=1}^{\infty}|ca_i|\,p_i=|c|\sum_{i=1}^{\infty}|a_i|p_i,$$

因为 $E\xi$ 存在而收敛，故 $E(c\xi)$ 存在.

$$E(c\xi)=\sum_{i=1}^{\infty}(ca_i)p_i=c\sum_{i=1}^{\infty}a_ip_i=cE\xi.$$

若 ξ 是连续型随机变量，有概率分布密度 $p(x)$. 这时积分

$$\int_{-\infty}^{+\infty}|cx|p(x)\,dx=|c|\int_{-\infty}^{+\infty}|x|p(x)dx,$$

因为 $E\xi$ 存在而收敛，故 $E(c\xi)$ 存在.

$$E(c\xi)=\int_{-\infty}^{+\infty}(cx)p(x)\,dx=c\int_{-\infty}^{+\infty}xp(x)\,dx=cE\xi. \quad ▌$$

以下的性质 3～6 将只对连续型的情形给出证明，离散型情形的证明留给读者.

性质 3 若随机向量 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的数学期望存在，则 $\xi_1+\xi_2+\cdots+\xi_n$ 的数学期望存在，并且

$$E(\xi_1+\xi_2+\cdots+\xi_n)=E\xi_1+E\xi_2+\cdots+E\xi_n.$$

【证】 设 $(\xi_1, \cdots, \xi_n)$ 有联合概率分布密度 $p(x_1, \cdots, x_n)$，这时 ξ_i 有概率分布密度

$$p_{\xi_i}(x_i)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p(x_1, \cdots, x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n,$$

$$i=1, 2, \cdots, n.$$

$$\begin{aligned}&\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}|x_1+\cdots+x_n|p(x_1, \cdots, x_n)dx_1\cdots dx_n\\ &\leqslant\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}|x_i|p(x_1, \cdots, x_n)dx_1\cdots dx_n\\ &=\sum_{i=1}^{n}\int_{-\infty}^{+\infty}|x_i|p_{\xi_i}(x_i)dx_i.\end{aligned}$$

因为 $E\xi_i(i=1, 2, \cdots, n)$ 存在，不等式右端是有限数，所以不等式左端的积分收敛，由 1.1 定理知 $E(\xi_1+\cdots+\xi_n)$ 存在.

$$\begin{aligned}&E(\xi_1+\cdots+\xi_n)\\ &=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(x_1+\cdots+x_n)p(x_1, \cdots, x_n)dx_1\cdots dx_n\\ &=\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}x_ip(x_1, \cdots, x_n)dx_1\cdots dx_n\\ &=\sum_{i=1}^{n}\int_{-\infty}^{+\infty}x_ip_{\xi_i}(x_i)dx_i=\sum_{i=1}^{n}E\xi_i. \quad ▌\end{aligned}$$

推论1 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)'$是$n$维随机向量，$l=(l_1, l_2, \cdots, l_n)$是任意给定的$n$维向量，则

$$El'\xi=l'E\xi.$$

【证】
$$El'\xi=E(l_1\xi_1+\cdots+l_n\xi_n)$$
$$=l_1E\xi_1+\cdots+l_nE\xi_n$$
$$=l'E\xi.\ \blacksquare$$

推论2 $\xi=(\xi_{ij})$是$m\times n$随机矩阵，A是任意给定的$l\times m$矩阵，则

$$EA\xi=AE\xi.$$

【证】设A的第i行为$a'_{(i)}$, $i=1, 2, \cdots, l$, ξ的第j列为ξ_j, $j=1, 2, \cdots n$，则

$$EA\xi=E\begin{pmatrix} a'_{(1)} \\ \vdots \\ a'_{(l)} \end{pmatrix}(\xi_1, \cdots, \xi_n)=E(a'_{(i)}\xi_j)=(Ea'_{(i)}\xi_j)$$

$$=(a'_{(i)}E\xi_j)=\begin{pmatrix} a'_{(1)} \\ \vdots \\ a'_{(l)} \end{pmatrix}(E\xi_1, \cdots, E\xi_n)=AE\xi.\ \blacksquare$$

性质4 随机变量ξ、η独立，若$E\xi$、$E\eta$存在，则$E(\xi\eta)$存在，并且

$$E(\xi\eta)=(E\xi)(E\eta).$$

【证】设(ξ, η)有概率分布密度$p(x, y)$.

这时，ξ、η分别有概率分布密度

$$p_\xi(x)=\int_{-\infty}^{+\infty}p(x, y)dy,$$

$$p_\eta(y)=\int_{-\infty}^{+\infty}p(x, y)dx,$$

并且

$$p(x, y)=p_\xi(x)\cdot p_\eta(y),$$

积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|xy|p(x, y)dxdy$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|x||y|p_\xi(x)p_\eta(y)dxdy$$

$$=\int_{-\infty}^{+\infty}|x|p_\xi(x)dx\cdot\int_{-\infty}^{+\infty}|y|p_\eta(y)dy<+\infty.$$

由1.1定理知 $E(\xi\eta)$ 存在，并且

$$E(\xi\eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyp(x, y)dxdy$$

$$=\int_{-\infty}^{+\infty}xp_\xi(x)dx\cdot\int_{-\infty}^{+\infty}yp_\eta(y)dy$$

$$=(E\xi)(E\eta).\quad \blacksquare$$

利用数学期望的性质有可能使一些随机变量数学期望的计算得到简化.

【例9】 在 n 次独立试验中每次试验成功的概率是 p. 设 η_i 表示第 i 次试验中成功的次数，则 η_i 有分布列

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix},$$

并且 $E\eta_i=0\times(1-p)+1\times p=p,\quad i=1, 2, \cdots, n.$

设 $$\eta=\eta_1+\eta_2+\cdots+\eta_n,$$

由性质3

$$E\eta=E(\eta_1+\cdots+\eta_n)=E\eta_1+\cdots+E\eta_n=np.$$

从定义容易看出，随机变量的数学期望由它的概率分布完全决定. 因此，具有相同分布的随机变量，必有相同的数学期望.

这里，η 服从二项分布(参数为 n, p)，从而若 ξ 是任一服从二项分布(参数为 n, p)的随机变量，则有

$$E\xi=E\eta=np.$$

在例 4 中已经直接求得了服从二项分布的随机变量的数学期望. 比较起来，本例使用的方法更为简洁一些. 现在再看一个类似的例子.

【例 10】 设 ξ 是自由度为 n 的 χ^2 变量. 在例 5 中已经求出了它的数学期望 $E\xi$. 现在我们使用另一方法求 $E\xi$.

设 $\eta_1, \cdots, \eta_n$ 是 n 个独立的标准正态随机变量. 我们知道，它们的平方和

$$\eta=\eta_1^2+\cdots+\eta_n^2$$

是自由度为 n 的 χ^2 变量. 由例 6,

$$E\eta_i^2=1, \quad i=1, \cdots, n.$$

由于 η 和 ξ 有相同分布，并利用性质 3，即得

$$E\xi=E\eta=E\eta_1^2+\cdots+E\eta_n^2=n.$$

【例 11】 为了保障人民身体健康，做到早期发现和预防某些疾病，经常要进行各种项目的普查. 有不少项目是对血液进行多种化学检验. 为了加快普查工作的进行，必须努力减少化验次数. 下面以对 1000 人的血液进行某项检验为例，介绍一种方法.

普查 1000 个人的血液，一般有两种方法：

(1) 对每个人的血液分别进行这项检验，如果血液呈阴性反应就断定受验者没有这种病；如果呈阳性，就断定有这种病. 这样就需要化验 1000 次.

(2) 如果在这项检验下，各人的血液混合起来没有交互作用，我们可以把 1000 个人随机地分成若干个组，每组 k 个人，把这 k 个人每人的血液取出一半混在一起进行化验. 如果这混合血液呈阴性，就说明每个人的血液都呈阴性，这样，对这 k 个人只要做一次化验，就能断定他们都没有这种病；如果呈阳性，那就说明这 k 个人中至少有一个人的血液呈阳性，

这时对每个人余下的一半血液逐个再化验一次以确定哪些人是患者.在这种情况下,对这 k 个人共做 $k+1$ 次化验.

当这种疾病的发病率比较低(即血液呈阳性率比较低)时,方法(2)能节约不少化验次数. 下面就来说明一下 k 取多少才能节约化验次数. 假设检验一个人的血液呈阳性的概率为 p, 呈阴性的概率为 $q=1-p$. 以 k 个人为一组时, 这 k 个人共需化验的次数 ξ 是一个随机变量, 按照方法(2)的规定, ξ 可能取的值有两个: 混合血液呈阴性时, ξ 取 1; 呈阳性时, ξ 取 $k+1$. 因为"$\xi=1$"表示"组内 k 个人每个人的血液都是阴性"这样一个事件, 又由于每个人的血液情况是互相独立的,所以

$$P(\xi=1)=qq\cdots q=q^k,$$
$$P(\xi=k+1)=1-q^k.$$

按照方法(2), 每 k 个人(即每组)的平均化验次数是

$$E\xi=1\times q^k+(k+1)\times(1-q^k)=k+1-kq^k.$$

根据性质 3, $\frac{1000}{k}$ 个组的平均化验次数是

$$\frac{1000}{k}(k+1-kq^k)=1000\left(1+\frac{1}{k}-q^k\right).$$

而按照方法(1), 1000 个人化验的次数是 1000 次. 若要使方法(2)比方法(1)减少化验次数就要使

$$1000\left(1+\frac{1}{k}-q^k\right)<1000,$$

即 $\frac{1}{k}<q^k$. 当 q 为已知数时,可以从这个不等式解出 k, 例如当 $p=0.3\%$, $q=99.7\%$ 时,解不等式

$$\frac{1}{k}<0.997^k.$$

借助对数表可以知道，当 $k\geqslant2$ 时上式都能成立．这就告诉我们，每组人数不小于2时，方法(2)都比方法(1)节省化验次数．

下面准备讨论一些关于数学期望的不等式，这将涉及到关于随机变量的不等式．设 $(\xi(\omega),\ \eta(\omega))$ 是定义在概率空间 $(\Omega,\ \mathscr{F},\ P)$ 上的二维随机变量，$\xi\leqslant\eta$ 表示对任何 $\omega\in\Omega$ 都有 $\xi(\omega)\leqslant\eta(\omega)$，也就是在每次试验中 ξ 取的值都不超过 η 取的值．因此当 $(\xi,\ \eta)$ 是连续型随机变量时，有一个概率分布密度 $p(x,\ y)$ 满足

$$p(x,\ y)=0,\quad x>y.$$

当 $(\xi,\ \eta)$ 是离散型随机变量，有分布列

$$P(\xi=x_i,\ \eta=y_j)=p_{ij},\quad i,\ j=1,\ 2,\ \cdots$$

时，对应于 $p_{ij}>0$ 的 $x_i,\ y_j$ 必满足 $x_i\leqslant y_j$．

性质5 若随机变量 ξ 只取非负值，并且 $E\xi$ 存在，则

$$E\xi\geqslant0.$$

【证】 设 ξ 是连续型随机变量，则有一个概率分布密度 $p(x)$ 满足 $p(x)=0,\ x\leqslant0$，所以

$$E\xi=\int_0^{+\infty}xp(x)dx\geqslant0.\ \blacksquare$$

由性质5不难得到

推论3 $$|E\xi|\leqslant E|\xi|.$$

证明留给读者．

性质6 $\xi,\ \eta$ 是二维随机变量，满足条件

$$0\leqslant\xi\leqslant\eta,$$

若 $E\eta$ 存在，则 $E\xi$ 存在，并且 $E\xi\leqslant E\eta$．

【证】 若 $(\xi,\ \eta)$ 是连续型随机变量，则有一个概率分布密度 $p(x,\ y)$，当 $x,\ y$ 不满足 $0\leqslant x\leqslant y$ 时，

$$p(x, y)=0.$$

故
$$\int_{-\infty}^{+\infty}|x|\left[\int_{-\infty}^{+\infty}p(x, y)dy\right]dx$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|x|p(x, y)dxdy$$
$$\leqslant\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|y|p(x, y)dxdy$$
$$=\int_{-\infty}^{+\infty}|y|\left[\int_{-\infty}^{+\infty}p(x, y)dx\right]dy<+\infty,$$

所以 $E\xi$ 存在.

因为 $$E\eta-E\xi=E(\eta-\xi)\geqslant 0,$$

所以 $$E\xi\leqslant E\eta.\quad \rrbracket$$

推论 4 (ξ, η) 是二维随机变量. 若 $|\xi|\leqslant c$(c 是常数), $E\eta$ 存在, 则 $E(\xi\eta)$ 存在, 并且

$$E|\xi\eta|\leqslant cE|\eta|.$$

性质 7 若 ξ 的 k 阶矩存在, 则对于任何小于 k 的非负整数 r, ξ 的 r 阶矩存在.

【证】 $E\xi^k$ 存在, 因而 $E(|\xi|^k+1)$ 存在. 由于

$$|\xi|^r\leqslant|\xi|^k+1,$$

由性质 6, $E|\xi|^r$ 存在, 因而 $E\xi^r$ 存在. $\rrbracket$

性质 8 若 $\xi\geqslant 0$, $E\xi$ 存在, 则对任何 $\varepsilon>0$ 有

$$P(\xi\geqslant\varepsilon)\leqslant\frac{E\xi}{\varepsilon}.$$

【证】 令

$$\eta=\begin{cases}\varepsilon, & \text{当 } \xi\geqslant\varepsilon \text{ 时},\\ 0, & \text{当 } \xi<\varepsilon \text{ 时},\end{cases}$$

则有 $0\leqslant\eta\leqslant\xi$. 由性质 6

$$\varepsilon P(\xi\geqslant\varepsilon)=E\eta\leqslant E\xi.\quad \rrbracket$$

性质 9 $\xi\geqslant 0$. 若 $E\xi=0$, 则 $P(\xi=0)=1$.

【证】 由性质 8, 对任何自然数 n 有

$$P\left(\xi\geqslant\frac{1}{n}\right)\leqslant nE\xi=0.$$

所以 $P(\xi>0)=P\left[\bigcup_{n=1}^{\infty}\left(\xi\geqslant\frac{1}{n}\right)\right]=\lim_{n\to\infty}P\left(\xi\geqslant\frac{1}{n}\right)=0,$

$$P(\xi=0)=1-P(\xi>0)=1. \quad ▌$$

性质 10 若 $E\xi^2$、$E\eta^2$ 存在,则 $E\xi\eta$ 存在,并且

$$(E\xi\eta)^2\leqslant E\xi^2\cdot E\eta^2. \tag{7}$$

这个不等式称为柯西-许瓦兹(Cauchy-Schwarz)不等式.

【证】 由于 $E(\xi^2+\eta^2)$ 存在,并且

$$|\xi\eta|\leqslant\xi^2+\eta^2,$$

所以 $E\xi\eta$ 存在.

当 $E\xi^2=0$ 时,

$$P(\xi\eta=0)\geqslant P(\xi=0)=1.$$

所以 $$P(\xi\eta=0)=1, \quad E\xi\eta=0,$$

(7)式成立.

当 $E\xi^2>0$ 时,对任何实数 t 有

$$t^2E\xi^2+2tE\xi+E\eta^2=E(t\xi+\eta)^2\geqslant0.$$

这说明 t 的二次方程

$$t^2E\xi^2+2tE\xi\eta+E\eta^2=0$$

或是没有实根,或是有重根. 因此它的判别式

$$(2E\xi\eta)^2-4E\xi^2\cdot E\eta^2\leqslant0.$$

这就证明了(7). ▌

以上我们对离散型随机变量和连续型随机变量这两种情形分别证明了数学期望的性质, 这些性质对于一般随机变量的数学期望也成立. 读者可以参看严士健等编《测度论与概率论基础》.

习 题 二

1. ξ是随机变量，$\nu_k=E\xi^k$，$\mu_k=E(\xi-E\xi)^k$，$k=1, 2, 3, 4$，试证

   (1) $\mu_3=\nu_3-3\nu_2\nu_1+2\nu_1^3$;

   (2) $\mu_4=\nu_4-4\nu_3\nu_1+6\nu_2\nu_1^2-3\nu_1^4$.

2. $\mu_k=E(\xi-E\xi)^k$，$k=2, 3, 4$. 称 $\gamma_1=\dfrac{\mu_3}{\mu_2^{\frac{3}{2}}}$ 为ξ分布的偏度，$\gamma_2=\dfrac{\mu_4}{\mu_2^2}-3$ 为ξ分布的峰度，试对下面分布求 γ_1, γ_2:

   (1) $N(0, 1)$;

   (2) 自由度为n的χ^2分布.

3. 设随机变量ξ, η独立，有相同的分布$N(0, 1)$，试证

$$E\max(\xi, \eta)=\frac{1}{\sqrt{\pi}}.$$

4. 设随机变量ξ, η独立，有相同的分布$N(a, \sigma^2)$，试求$E\max(\xi, \eta)$.

5. 证明

   (1) $|E\xi|\leqslant E|\xi|$;

   (2) 若$|\eta|\leqslant c$(c是常数)，则$|E\xi\eta|\leqslant cE|\xi|$.

第二节 方 差

2.1 方差的定义

在随机变量的数字特征中，一阶矩、二阶矩比较重要. 这一节将对随机变量的二阶中心矩做一些讨论.

称随机变量ξ的二阶中心矩$E(\xi-E\xi)^2$为ξ的**方差**，记作$D\xi$. 根据数学期望的定义，当ξ有分布列$P(\xi=x_i)=p_i$，$i=1, 2, \cdots$时，

$$D\xi=\sum_{i=1}^{\infty}(x_i-E\xi)^2p_i. \tag{1}$$

当ξ有概率分布密度$p(x)$时，

$$D\xi=\int_{-\infty}^{+\infty}(x-E\xi)^2p(x)dx. \qquad (2)$$

在计算 $D\xi$ 时，有时用下面的公式比较方便：

$$D\xi=E\xi^2-(E\xi)^2. \qquad (3)$$

事实上，由 1.2 性质 3 知

$$\begin{aligned}D\xi&=E(\xi-E\xi)^2\\&=E[\xi^2-2\xi(E\xi)+(E\xi)^2]\\&=E\xi^2-2(E\xi)^2+(E\xi)^2\\&=E\xi^2-(E\xi)^2.\end{aligned}$$

方差的含义是什么呢？我们在上一节讲数学期望时曾说过，一个随机变量的数学期望是一个常数，它在一定意义下表示了这个随机变量的平均值. 正因为这样，当一个随机变量取值时，取到的值总会有的比 $E\xi$ 大，有的比 $E\xi$ 小. 也就是说，ξ 总是在 $E\xi$ 的周围取值. 但是在 $E\xi$ 周围取值的情况也有不同，有的随机变量取的值密集在它的数学期望周围，有的则比较分散. 图 3-1, 3-2 是这两种情况的示意图.

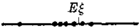


图 3-1

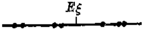


图 3-2

为了区别这些不同的情况，需要定义一个刻划随机变量取值分散程度(反过来说是集中程度)的数. 因为随机变量 ξ 的方差是 ξ 取的值与 $E\xi$ 距离平方的平均值，若 ξ 取的值比较集中，则 $D\xi$ 小；反之，若 ξ 取的值比较分散，则 $D\xi$ 大，所以 $D\xi$ 是刻划 ξ 取值分散程度的一个数. 它是 ξ 的一个重要数字特征.

【例 1】 $\xi\sim N(a,\sigma^2)$，试求 $D\xi$.

解：

$$D\xi=\int_{-\infty}^{+\infty}(x-E\xi)^2\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{+\infty}(x-a)^2\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(\frac{x-a}{\sigma}\right)^2e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}d\left(\frac{x-a}{\sigma}\right).$$

由第一节例 6, $D\xi=\sigma^2$.

在第二章讲正态随机变量时曾经解释过参数 σ^2 的含义，说 σ^2 刻划了随机变量取值的分散程度. 现在更明确了，σ^2 就是 ξ 的方差.

【例 2】 ξ 服从参数为 λ 的泊松分布，试求 $D\xi$.

解：由第一节例 1 知 $E\xi=\lambda$, 这里再计算 $E\xi^2$, 然后利用(3)式计算 $D\xi$.

$$E\xi^2=\sum_{k=0}^{+\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}$$

$$=\sum_{k=0}^{+\infty}(k^2-k)\frac{\lambda^k}{k!}e^{-\lambda}+\sum_{k=0}^{+\infty}k\frac{\lambda^k}{k!}e^{-\lambda}$$

$$=\sum_{k=2}^{+\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}+\lambda=\lambda^2\sum_{k-2=0}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}+\lambda$$

$$=\lambda^2+\lambda.$$

所以 $$D\xi=E\xi^2-(E\xi)^2=(\lambda^2+\lambda)-\lambda^2=\lambda.$$

这里我们看到了泊松分布中的参数 λ 既是分布的数学期望，也是分布的方差.

【例 3】 ξ 服从参数为 (n, p) 的二项分布，试求 $D\xi$.

解：$n=1$ 时，ξ^2 有分布列

$$P(\xi^2=1)=p,\quad P(\xi^2=0)=1-p,\quad E\xi^2=p,$$

$$D\xi=E\xi^2-(E\xi)^2=p-p^2=p(1-p).$$

$n\geqslant 2$ 时，

$$
\begin{aligned}
E\xi^2 &= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n} (k^2-k) \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}+np \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} \\
&\quad \times p^{k-2}(1-p)^{n-k}+np \\
&= n(n-1)p^2[p+(1-p)]^{n-2}+np \\
&= n^2p^2-np^2+np,
\end{aligned}
$$

$$
\begin{aligned}
D\xi &= E\xi^2-(E\xi)^2=(n^2p^2-np^2+np)-(np)^2 \\
&= np(1-p).
\end{aligned}
$$

【例 4】 膨胀仪是测量金属膨胀系数的一种精密仪器，测量结果通过感光设备在照相底片上显示出来．现在同一台膨胀仪上，用两种底片多次测量某种合金的膨胀系数，结果列成下表：

玻璃底片测量结果

测量结果 ξ_1	13.4	13.5	13.6	13.7	13.8
概　率 P_{ξ_1}	0.05	0.15	0.60	0.15	0.05

软片测量结果

测量结果 ξ_2	13.3	13.4	13.5	13.6	13.7	13.8	13.9
概　率 P_{ξ_2}	0.05	0.05	0.15	0.50	0.15	0.05	0.05

试问哪一种底片的效果较好？

解： 比较 ξ_1 和 ξ_2 的方差，看用哪一种方法得到的测量结果更集中．测量结果比较集中的表明效果好．为了求 ξ_1 和 ξ_2 的方差，先分别求它们的数学期望：

$$
\begin{aligned}
E\xi_1 &= 13.4\times0.05+13.5\times0.15+13.6\\
&\quad\times0.60+13.7\times0.15+13.8\times0.05\\
&=13.6,\\
E\xi_2 &= 13.3\times0.05+13.4\times0.05+13.5\\
&\quad\times0.15+13.6\times0.50+13.7\times0.15+13.8\\
&\quad\times0.05+13.9\times0.05\\
&=13.6,\\
D\xi_1 &= (13.4-13.6)^2\times0.05+(13.5-13.6)^2\\
&\quad\times0.15+(13.6-13.6)^2\times0.60\\
&\quad+(13.7-13.6)^2\times0.15\\
&\quad+(13.8-13.6)^2\times0.05\\
&=0.0070,\\
D\xi_2 &= (13.3-13.6)^2\times0.05+(13.4-13.6)^2\\
&\quad\times0.05+(13.5-13.6)^2\times0.15\\
&\quad+(13.6-13.6)^2\times0.50\\
&\quad+(13.7-13.6)^2\times0.15\\
&\quad+(13.8-13.6)^2\times0.05\\
&\quad+(13.9-13.6)^2\times0.05\\
&=0.0160.
\end{aligned}
$$

因为 $D\xi_1<D\xi_2$，故用玻璃底片的效果较好。

2.2 方差的性质

在下列性质中，假定涉及到的随机变量的方差都存在*)。

性质 1 $Dc=0$ （c 是任意常数）。

【证】 $Dc=E(c-Ec)^2=E(c-c)^2=0$. ▌

*) 性质 2, 3, 4 都可以叙述为：若等式右端的方差存在，则左端方差存在且等式成立。读者可以用方差的定义给以证明。

性质 2　$D(c\xi)=c^2D\xi$　(c是任意常数).

【证】

$$D(c\xi)=E[c\xi-E(c\xi)]^2=E[c\xi-cE\xi]^2$$
$$=E[c^2(\xi-E\xi)^2]=c^2E(\xi-E\xi)^2=c^2D\xi.\quad ▌$$

性质 3　$D(\xi+c)=D\xi$　(c是任意常数).

【证】　$D(\xi+c)=E[(\xi+c)-E(\xi+c)]^2$
$$=E(\xi-E\xi)^2=D\xi.\quad ▌$$

性质 4　若 $\xi_1, \xi_2, \cdots, \xi_n$ 两两独立, 则
$$D(\xi_1+\cdots+\xi_n)=D\xi_1+\cdots+D\xi_n.$$

【证】

$$D(\xi_1+\cdots+\xi_n)=E\left[\sum_{i=1}^{n}\xi_i-E\left(\sum_{i=1}^{n}\xi_i\right)\right]^2$$
$$=E\left[\sum_{i=1}^{n}(\xi_i-E\xi_i)\right]^2$$
$$=E\left[\sum_{i=1}^{n}(\xi_i-E\xi_i)^2+2\sum\sum_{1\leq i<j\leq n}(\xi_i-E\xi_i)(\xi_j-E\xi_j)\right].$$

因为 $i\neq j$ 时, ξ_i, ξ_j 独立, $E(\xi_i\xi_j)=(E\xi_i)(E\xi_j)$,
$$E(\xi_i-E\xi_i)(\xi_j-E\xi_j)$$
$$=E(\xi_i-E\xi_i)\xi_j-E(\xi_i-E\xi_i)E\xi_j$$
$$=E(\xi_i\xi_j)-(E\xi_i)(E\xi_j)-(E\xi_i-E\xi_i)E\xi_j$$
$$=0,$$

所以　$D(\xi_1+\cdots+\xi_n)=\sum_{i=1}^{n}E(\xi_i-E\xi_i)^2=\sum_{i=1}^{n}D\xi_i.$　▌

【例 5】 例 3 中设 ξ_i 表示 n 次独立试验中事件 A 在第 i 次试验出现的次数, 则 $\xi_1, \cdots, \xi_n$ 独立, ξ_i 有分布列
$$P(\xi_i=1)=p,\quad P(\xi_i=0)=1-p,$$
$$i=1, 2, \cdots, n.$$

它们的方差　$D\xi_i=p(1-p)$, $i=1, 2, \cdots, n.$

利用性质 4，

$$D(\xi_1+\cdots+\xi_n)=D\xi_1+\cdots+D\xi_n$$
$$=np(1-p).$$

$\xi_1+\cdots+\xi_n$ 是 n 次独立试验中事件 A 出现的次数，服从二项分布(参数为 $(n,\ p)$)． 具有相同分布的随机变量有相同的方差，所以，任一随机变量 ξ 服从二项分布(参数为 $(n,\ p)$)，则

$$D\xi=D(\xi_1+\cdots+\xi_n)=np(1-p).$$

在例 3 中我们已经直接求出了服从二项分布的随机变量的方差，比较起来，这里的方法更为简单一些．

【例 6】 试求自由度为 n 的 χ^2 变量 ξ 的方差．

解： 设 $\xi_1,\ \cdots,\ \xi_n$ 是 n 个独立的标准正态随机变量．对于每一 i，$1\leqslant i\leqslant n$，

$$D\xi_i^2=E\xi_i^4-(E\xi_i^2)^2$$
$$=E\xi_i^4-[D\xi_i+(E\xi_i)^2]^2$$
$$=E\xi_i^4-1.$$

由第一节例 6，$E\xi_i^4=3$，所以

$$D\xi_i^2=3-1=2.$$

再根据性质 4，

$$D(\xi_1^2+\cdots+\xi_n^2)=D\xi_1^2+\cdots+D\xi_n^2=2n.$$

$\xi_1^2+\cdots+\xi_n^2$ 是自由度为 n 的 χ^2 变量，与 ξ 有相同的分布，故

$$D\xi=D(\xi_1^2+\cdots+\xi_n^2)=2n.$$

【例 7】 $\xi_1,\ \xi_2,\ \cdots,\ \xi_n$ 独立．设 $E\xi_i=a$，$D\xi_i=\sigma^2$，$i=1,\ 2,\ \cdots,\ n$． 试求 $\bar{\xi}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ 的数学期望和方差．

解： 由数学期望和方差的性质有

$$E\bar{\xi}=E\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\xi_i=a,$$

$$D\bar{\xi}=D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n^2}D\left(\sum_{i=1}^{n}\xi_i\right)$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}D\xi_i=\frac{\sigma^2}{n}.$$

当我们进行精密测量时，为了减少随机误差，往往是重复测量几次然后取测量结果的平均值．例 7 给这种做法以一个合理的解释．设被测量的真值为 a，由于有随机误差，n 次测量的结果 $\xi_1, \cdots, \xi_n$ 是 n 个随机变量，数学期望都是 a．在测量条件保持不变，且每次测量都是"独立"进行的情况下，$\xi_1, \cdots, \xi_n$ 是独立的，有相同的分布．当 $n=1$ 时，测量结果 ξ_1 在真值 a 周围取值，方差为 $D\xi_1=\sigma^2$．当 $n>1$ 时，n 次测量结果的平均值 $\bar{\xi}$ 仍在真值 a 周围取值，但方差为 $\frac{\sigma^2}{n}$，比 σ^2 小，因此 $\bar{\xi}$ 更有可能取得接近于真值 a 的值．

2.3 车贝谢夫不等式

从这两节的例题可以看到：许多常用的随机变量的分布，在函数形式已知时完全由它的数学期望和方差所决定，如泊松分布，正态分布，二项分布，χ^2 分布等等．可见这两个数字特征是很重要的．不仅如此，它们的重要性还在于当分布的函数形式不知道时，也能提供关于分布的某些信息．这从下面著名的车贝谢夫(Чебышев)不等式可以看到．

车贝谢夫不等式：若随机变量 ξ 的方差 $D\xi$ 存在，则对任何 $\varepsilon>0$，有

$$P(|\xi-E\xi|\geqslant\varepsilon)\leqslant\frac{D\xi}{\varepsilon^2}.$$

【证】 由 1.2 性质 8,

$$P(|\xi-E\xi|\geqslant\varepsilon)=P(|\xi-E\xi|^2\geqslant\varepsilon^2)$$
$$\leqslant\frac{E|\xi-E\xi|^2}{\varepsilon^2}=\frac{D\xi}{\varepsilon^2}.\quad]$$

若取 $\varepsilon=3\sqrt{D\xi}$, 便有

$$P(|\xi-E\xi|\geqslant3\sqrt{D\xi})\leqslant\frac{D\xi}{(3\sqrt{D\xi})^2}=\frac{1}{9};$$

取 $\varepsilon=2\sqrt{D\xi}$, 便有

$$P(|\xi-E\xi|\geqslant2\sqrt{D\xi})\leqslant\frac{D\xi}{(2\sqrt{D\xi})^2}=\frac{1}{4}.$$

所以, 利用 $E\xi$、$D\xi$ 可以得到关于"$|\xi-E\xi|\geqslant k\sqrt{D\xi}$"这类事件概率的粗略估计.

习 题 三

1. 随机变量 ξ 在区间 $[a, b]$ 上均匀分布, 试求 $D\xi$. (结合这个习题再理解方差的含义)
2. 试求拉普拉斯分布的方差.
3. 试求马克斯威尔分布的方差.
4. (ξ, η) 服从参数为 $(a, b, \sigma_1^2, \sigma_2^2, \rho)$ 的二维正态分布, 试求 $D\xi$, $D\eta$.
5. $a\leqslant\xi\leqslant b$, 试证明: $a\leqslant E\xi\leqslant b$, $\quad D\xi\leqslant\left(\frac{b-a}{2}\right)^2$.
6. 证明 $f(C)=E(\xi-C)^2$ 在 $C=E\xi$ 处极小.
7. ξ 服从对数正态分布, 即有概率分布密度

$$p(x)=\begin{cases}0, & \text{当 } x\leqslant0;\\ \dfrac{1}{x\beta\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2\beta^2}(\ln x-a)^2\right\}, & \text{当 } x>0.\end{cases}$$

试求 $E\xi$, $D\xi$.

8. 证明 $D\xi=0$ 的充分必要条件是

$$P(\xi=E\xi)=1.$$

第三节　相关系数、协方差与协方差阵

3.1 回归直线

设 ξ, η 是二维随机变量，$D\xi>0$, $D\eta>0$. 我们希望找到一个线性函数 $y=\alpha+\beta x$, 使 $\alpha+\beta\xi$ 能尽量好地表示 η, 也就是希望找到 α、β, 使

$$E[\eta-(\alpha+\beta\xi)]^2=\min_{a,b} E[\eta-(a+b\xi)]^2. \tag{1}$$

为了确定 α, β, 展开(1)式的右端:

$$\begin{aligned}
&E[\eta-(a+b\xi)]^2\\
&=E[(\eta-E\eta)-b(\xi-E\xi)+(E\eta-a-bE\xi)]^2\\
&=b^2E(\xi-E\xi)^2-2bE(\xi-E\xi)(\eta-E\eta)\\
&\quad+E(\eta-E\eta)^2+(E\eta-a-bE\xi)^2\\
&\quad+2(E\eta-a-bE\xi)E(\eta-E\eta)\\
&\quad+2b(E\eta-a-bE\xi)E(\xi-E\xi)\\
&=b^2D\xi-2bE(\xi-E\xi)(\eta-E\eta)+D\eta\\
&\quad+(E\eta-a-bE\xi)^2.
\end{aligned} \tag{2}$$

最后一步利用了 $E(\xi-E\xi)=0$, $E(\eta-E\eta)=0$. (2)式右端的前三项是 b 的二次三项式，因为 b^2 的系数 $D\xi>0$, 二次三项式在 $\dfrac{E(\xi-E\xi)(\eta-E\eta)}{D\xi}$ 处达到极小，又(2)式右端第四项非负，故

$$\beta=\frac{E(\xi-E\xi)(\eta-E\eta)}{D\xi}, \tag{3}$$

$$\alpha=E\eta-\frac{E(\xi-E\xi)(\eta-E\eta)}{D\xi}\cdot E\xi. \tag{4}$$

称 $y=\alpha+\beta x$ 为 η 关于 ξ 的回归直线. 将(3)、(4)代入(2),

可以得到 η 与 $\alpha+\beta\xi$ 的平均平方距离

$$
\begin{aligned}
E[\eta-(\alpha+\beta\xi)]^2
&=\frac{[E(\xi-E\xi)(\eta-E\eta)]^2}{D\xi}\\
&\quad-2\frac{[E(\xi-E\xi)(\eta-E\eta)]^2}{D\xi}+D\eta\\
&=D\eta-\frac{[E(\xi-E\xi)(\eta-E\eta)]^2}{D\xi}.
\end{aligned}
\tag{5}
$$

3.2 相关系数

相关系数是刻划两个随机变量间线性相关程度的一个数值. 从 3.1 可以看到, 对任何二维随机变量 (ξ,η) $(D\xi>0)$, 都可以求出 η 关于 ξ 的回归直线, 用这条直线可以表示 ξ,η 间的线性关系. 然而, 两个随机变量之间的线性关系还有不同的情况. 图 3-3 给出了几个二维离散型随机变量的概率

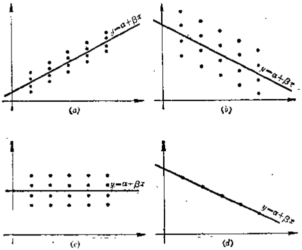


图 3-3

分布示意图，每个二维随机变量以相同的概率取标有“.”的点，$y=\alpha+\beta x$ 是回归直线.

当 $\beta>0$ 时，如图 3-3a，对应于 ξ 取的每一个值 x，虽然 η 取的值并不确定，而且有随机性，但随着 ξ 取值的增大，η 的取值有增大的趋势. 我们称这种现象为 η 对 ξ 有正的线性相关关系. $\beta<0$ 时，如图 3-3b)，随着 ξ 取值的增大，η 的取值有下降的趋势. 这时我们称 η 对 ξ 有负的线性相关关系. 图 3-3c 是 $\beta=0$ 的情形. 由 (4)，$\alpha=E\eta$，直线 $y=\alpha+\beta x=E\eta$ 平行于 x 轴，当 ξ 取的值增大或减小时，η 总是围绕它的数学期望 $E\eta$ 取值，没有任何增大或减小的趋势. 在这个意义上我们说 η 和 ξ 是没有线性相关关系的，或简单地说：η 和 ξ 线性无关.

在 η 和 ξ 有线性相关关系(正相关或负相关)时，如果(ξ, η)的概率分布相对于 $D\eta$ 比较集中在直线 $y=\alpha+\beta x$ 附近，就说 ξ, η 的线性相关程度比较高，这时

$$\frac{E[\eta-(\alpha+\beta\xi)]^2}{D\eta}=1-\frac{[E(\xi-E\xi)(\eta-E\eta)]^2}{D\xi\cdot D\eta} \qquad (6)$$

比较小；对相反的情形，就说 ξ, η 的线性相关程度比较低，这时 (6) 也比较大. 图 3-3d 中 η 是 ξ 的线性函数，$\eta=\alpha+\beta\xi$ $(\beta\neq0)$，(ξ, η)的概率分布完全集中在直线 $y=\alpha+\beta x$ 上，η 与 ξ 的线性相关程度达到最高. 图3-3a 中，η 与 ξ 的线性相关程度就不如图 3-3d 中的高，但比图 3-3b 中的高. 图 3-3c 中 η 和 ξ 没有线性相关关系，线性相关程度最低.

我们希望用一个数值来刻划ξ, η 线性相关关系的各种不同情况，希望 η 对 ξ 没有线性相关关系时数值为 0，有正线性相关关系时数值为正，有负线性相关关系时数值为负，相关程度越高，数值的绝对值越大，线性相关程度最高(即 η 是 ξ

的线性函数)时，数值的绝对值为1．数

$$\rho=\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi}\sqrt{D\eta}} \tag{7}$$

就满足上面的要求．

从(6)式知，

$$1-\frac{[E(\xi-E\xi)(\eta-E\eta)]^2}{D\xi\cdot D\eta}\geqslant 0,$$

所以 $\rho^2\leqslant 1$, $-1\leqslant\rho\leqslant 1$. 从(3)式知，$\rho$ 与 β 同号，所以有正相关关系时 ρ 为正，有负相关关系时 ρ 为负．又

$$\frac{E[\eta-(\alpha+\beta\xi)]^2}{D\xi}=1-\rho^2.$$

所以相关程度越高，ρ 的绝对值越大．由1.2性质9知，当且仅当 $P(\eta=\alpha+\beta\xi)=1$, 即 (ξ,η) 的概率分布集中在直线 $y=\alpha+\beta x$ 上时，$1-\rho^2=0$, $|\rho|=1$. ξ、η 线性无关时，$\beta=0$ 因而 $\rho=0$.

定义 称(7)式定义的 ρ 为 ξ、η 的**相关系数**．

在相关系数的定义中，ξ、η 是对称的．就是说，如果直线 $x=\alpha'+\beta'y$ 使

$$E[\xi-(\alpha'+\beta'\eta)]^2=\min_{a',b'}E[\xi-(a'+b'\eta)]^2,$$

也有

$$\frac{E[\xi-(\alpha'+\beta'\eta)]^2}{D\xi}=1-\rho^2.$$

读者不妨自己给以证明．

需要强调的是，二维随机变量的相关系数只是刻划它们之间线性相关程度的数值．$\rho=0$ 只表示 ξ, η 线性无关，并不是 ξ, η 没有什么关系．例如，考虑随机变量 (ξ,η), ξ 在 $[-1,1]$ 上均匀分布，$\eta=\xi^2$. 这时 ξ, η 之间是有关系的，但

$$E(\xi-E\xi)(\eta-E\eta)=E\xi(\xi^2-E\xi^2)$$
$$=E\xi^3-(E\xi^2)(E\xi)=0,$$

因而 $\rho=0$.

【例 1】 (ξ, η)是二维正态随机变量,有概率分布密度

$$p(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}$$
$$\times\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}\right.\right.$$
$$\left.\left.-2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\},$$

试求 ξ, η 的相关系数.

解:

$$E(\xi-E\xi)(\eta-E\eta)$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-a)(y-b)p(x, y)dx\,dy$$
$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-a)(y-b)$$
$$\times\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right)^2\right.$$
$$\left.-\frac{(y-b)^2}{2\sigma_2^2}\right\}dx\,dy.$$

取

$$z=\frac{1}{\sqrt{1-r^2}}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right),$$
$$t=\frac{y-b}{\sigma_2},$$

则

$$\frac{\partial(x, y)}{\partial(z, t)}=\sqrt{1-r^2}\sigma_1\sigma_2,$$

$$E(\xi-E\xi)(\eta-E\eta)$$
$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sigma_1(\sqrt{1-r^2}z+rt)\sigma_2 t$$
$$\times\exp\left\{-\frac{z^2}{2}-\frac{t^2}{2}\right\}dz\,dt$$
$$=\frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi}\int_{-\infty}^{+\infty}ze^{-\frac{z^2}{2}}dz$$
$$\times\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2}}dt+\frac{r\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz\int_{-\infty}^{+\infty}t^2e^{-\frac{t^2}{2}}dt$$
$$=r\sigma_1\sigma_2=r\sqrt{D\xi}\sqrt{D\eta},$$

故 $\rho=r$.

3.3 协方差、协方差阵

定义 ξ, η 是二维随机变量，若 $E(\xi-E\xi)(\eta-E\eta)$ 存在，就称它为 ξ, η 的*协方差*，记作 $\mathrm{cov}(\xi,\eta)$.

由于

$$|(\xi-E\xi)(\eta-E\eta)|\leqslant\frac{1}{2}[(\xi-E\xi)^2+(\eta-E\eta)^2].$$

当 $D\xi$, $D\eta$ 都存在时，$\mathrm{cov}(\xi,\eta)$ 一定存在. $\mathrm{cov}(\xi,\eta)$ 也是刻划 ξ, η 线性相关程度的数字特征，这是因为 ξ, η 的相关系数

$$\rho=\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$$
$$=\frac{\mathrm{cov}(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}},$$

在 $D\xi$, $D\eta$ 一定的情况下，$|\rho|$ 的大小取决于 $|\mathrm{cov}(\xi,\eta)|$ 的大小.

定义 $\xi'=(\xi_1,\cdots,\xi_m)$ 是 m 维随机向量，$\eta'=(\eta_1,\cdots,\eta_n)$ 是 n 维随机向量. 若对任何 i, j $(1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n)$, $\mathrm{cov}(\xi_i,\eta_j)$ 存在，就称矩阵

$$\begin{pmatrix} \text{cov}(\xi_1, \eta_1) & \text{cov}(\xi_1, \eta_2) & \cdots & \text{cov}(\xi_1, \eta_n) \\ \text{cov}(\xi_2, \eta_1) & \text{cov}(\xi_2, \eta_2) & \cdots & \text{cov}(\xi_2, \eta_n) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}(\xi_m, \eta_1) & \text{cov}(\xi_m, \eta_2) & \cdots & \text{cov}(\xi_m, \eta_n) \end{pmatrix}$$

为ξ, η的协方差阵，记作$C(\xi, \eta)$. $C(\xi, \eta)$的向量表达式是

$$C(\xi, \eta)=E(\xi-E\xi)(\eta-E\eta)',$$

这是因为

$$\begin{aligned} C(\xi, \eta) &= (\text{cov}(\xi_i, \eta_j)) = (E(\xi_i-E\xi_i)(\eta_j-E\eta_j)) \\ &= E((\xi_i-E\xi_i)(\eta_j-E\eta_j)) \\ &= E\begin{pmatrix} \xi_1-E\xi_1 \\ \xi_2-E\xi_2 \\ \vdots \\ \xi_m-E\xi_m \end{pmatrix}(\eta_1-E\eta_1, \eta_2-E\eta_2, \cdots, \eta_n-E\eta_n) \\ &= E(\xi-E\xi)(\eta-E\eta)'. \end{aligned}$$

特别，当$\xi=\eta$时，称协方差阵$C(\xi, \xi)$为ξ的协方差阵，记作$D(\xi)$.

协方差与协方差阵具有下列性质：

设$\xi=(\xi_1, \xi_2, \cdots, \xi_m)'$, $\eta=(\eta_1, \eta_2, \cdots, \eta_n)'$是随机向量.

性质1 $C(\xi, \eta)$的转置是$C(\eta, \xi)$.

性质2 若a是列向量，

$$P(\eta=a)=1, \quad 则 \quad C(\xi, \eta)=0.$$

性质3 若ξ, η独立，则

$$C(\xi, \eta)=0.$$

【证】 对任何i, $j(1\leqslant i\leqslant m, 1\leqslant j\leqslant n)$，由于$\xi$, η独立，因而有ξ_i, η_j独立，所以$\text{cov}(\xi_i, \eta_j)=0$,

$$C(\xi, \eta)=(\text{cov}(\xi_i, \eta_j))=0. \quad \rbrack$$

性质 4 对任何实向量

$$a=(a_1, a_2, \cdots, a_m)', \quad b=(b_1, b_2, \cdots, b_n)',$$
$$\mathrm{cov}(a'\xi, b'\eta)=a'C(\xi, \eta)b.$$

【证】

$$\begin{aligned}
&\mathrm{cov}(a'\xi, b'\eta)\\
&=E\left[\sum_{i=1}^{m}a_i\xi_i-E\left(\sum_{i=1}^{m}a_i\xi_i\right)\right]\\
&\quad\times\left[\sum_{j=1}^{n}b_j\eta_j-E\left(\sum_{j=1}^{n}b_j\eta_j\right)\right]\\
&=E\left[\sum_{i=1}^{m}a_i(\xi_i-E\xi_i)\sum_{j=1}^{n}b_j(\eta_j-E\eta_j)\right]\\
&=\sum_{i=1}^{m}\sum_{j=1}^{n}a_ib_jE(\xi_i-E\xi_i)(\eta_j-E\eta_j)\\
&=\sum_{i=1}^{m}\sum_{j=1}^{n}a_ib_j\,\mathrm{cov}(\xi_i, \eta_j)\\
&=(a_1, \cdots, a_m)(\mathrm{cov}(\xi_i, \eta_j))\begin{pmatrix}b_1\\ \vdots\\ b_n\end{pmatrix}\\
&=a'C(\xi, \eta)b.
\end{aligned}$$
】

性质 5 对任何 $p\times m$ 矩阵 A, $q\times n$ 矩阵 B

$$C=(A\xi, B\eta)=AC(\xi, \eta)B'.$$

【证】 设*)

$$A=\begin{pmatrix}a'_{(1)}\\ \vdots\\ a'_{(p)}\end{pmatrix}, \qquad B=\begin{pmatrix}b'_{(1)}\\ \vdots\\ b'_{(q)}\end{pmatrix},$$

$a'_{(i)}(i=1, 2, \cdots, p)$是 A 的行向量, $b'_{(j)}(j=1, 2, \cdots, q)$是 B 的行向量,则

*) 本书用到的关于矩阵分块的运算可参看谢邦杰《线性代数》。

$$C(A\xi,\ B\eta)=C\left(\begin{pmatrix} a'_{(1)}\xi \\ \vdots \\ a'_{(p)}\xi \end{pmatrix},\ \begin{pmatrix} b'_{(1)}\eta \\ \vdots \\ b'_{(q)}\eta \end{pmatrix}\right)$$

$$=(\operatorname{cov}(a'_{(i)}\xi,\ b'_{(j)}\eta))=(a'_{(i)}C(\xi,\ \eta)b_{(j)})$$

$$=\begin{pmatrix} a'_{(1)} \\ \vdots \\ a'_{(p)} \end{pmatrix}C(\xi,\ \eta)(b_{(1)},\ \cdots,\ b_{(q)})$$

$$=AC(\xi,\ \eta)B'. \quad \rbrack$$

特别,若 $\xi_1,\ \xi_2,\ \cdots,\ \xi_r$ 都是 m 维随机向量, $\eta_1,\ \eta_2,\ \cdots,\ \eta_s$ 都是 n 维随机向量时,

$$C\left(\sum_{i=1}^{r}\xi_i,\ \sum_{j=1}^{s}\eta_j\right)$$

$$=C\left((I_m,\ \cdots,\ I_m)\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_r \end{pmatrix},\ (I_n,\ \cdots,\ I_n)\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_s \end{pmatrix}\right)$$

$$=(I_m,\ \cdots,\ I_m)(C(\xi_i,\ \eta_j))\begin{pmatrix} I_n \\ \vdots \\ I_n \end{pmatrix}$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}C(\xi_i,\ \eta_j),$$

其中 I_m 表示 m 阶单位方阵.

性质 6 $\xi=(\xi_1,\ \cdots,\ \xi_m)$ 是随机向量, $D(\xi)$ 存在, 则 $D(\xi)$ 是非负定方阵.

【证】 $D(\xi)=(\operatorname{cov}(\xi_i,\ \xi_j))$. 因为对任何 $i,\ j\ (1\leqslant i,\ j\leqslant m)$,

$$\operatorname{cov}(\xi_i,\ \xi_j)=E(\xi_i-E\xi_i)(\xi_j-E\xi_j)$$
$$=\operatorname{cov}(\xi_j,\ \xi_i),$$

所以 $D(\xi)$ 是对称方阵. 又对任何实向量 $x'=(x_1,\ \cdots,\ x_m)$,

$$x'D(\xi)x = x'C(\xi,\ \xi)x = \operatorname{cov}(x'\xi,\ x'\xi)$$
$$= D(x'\xi) \geqslant 0,$$

所以 $D(\xi)$ 是非负定矩阵. ▌

【例 2】 $\xi=(\xi_1,\ \xi_2,\ \cdots,\ \xi_m)'$ 服从参数为 a、B 的多维正态分布，a 是 m 维列向量，B 是 m 阶正定阵，试求 ξ 的协方差阵.

解： 设 $\eta_1,\ \cdots,\ \eta_m$ 是 m 个独立标准正态随机变量，则 $\eta=(\eta_1,\ \cdots,\ \eta_m)'$ 的协方差阵为 $D(\eta)=I_m$，I_m 是 m 阶单位阵. B 是正定阵，存在满秩方阵 C 使 $B=CC'$. 考虑随机向量 $C\eta+a$，由第二章第六节例 8 知，$C\eta+a$ 服从参数为 a、B 的多维正态分布，即与 ξ 有相同的分布，而服从相同分布的随机向量有相同的协方差阵，所以

$$D(\xi)=D(C\eta+a)=D(C\eta)$$
$$=CD(\eta)C'=CI_mC'$$
$$=CC'=B.$$

这里我们看到多维正态分布中正定阵 B 的含义，它就是分布的协方差阵.

在性质 3 中证明了 ξ, η 独立时，$C(\xi,\ \eta)$ 必为零. 逆命题不一定对，在 3.2 中已有反例. 但对正态随机变量，这逆命题却是成立的.

性质 7 (i) 设 $(\xi_1,\ \cdots,\ \xi_m)'$ 是 m 维正态随机向量，则 $\xi_1,\ \cdots,\ \xi_m$ 独立的充要条件是

$$\operatorname{cov}(\xi_i,\ \xi_j)=0,\quad i\neq j,\quad 1\leqslant i,\ j\leqslant m.$$

(ii) 设 $\xi=(\xi_1,\ \cdots,\ \xi_m)'$，$\eta=(\eta_1,\ \cdots,\ \eta_n)'$，$(\xi',\ \eta')'=(\xi_1,\ \cdots,\ \xi_m,\ \eta_1,\ \cdots,\ \eta_n)'$ 服从 $m+n$ 维正态分布，则 ξ, η 独立的充要条件是 $C(\xi,\ \eta)=0$.

【证】 (i) 只需证明条件的充分性.

假定条件成立，

$$\operatorname{cov}(\xi_i, \xi_j)=0, \quad i\neq j.$$

设$(\xi_1, \cdots, \xi_m)'\sim N(a, B)$，由例2知

$$B=\begin{pmatrix} D(\xi_1) & \operatorname{cov}(\xi_1, \xi_2) & \cdots & \operatorname{cov}(\xi_1, \xi_m) \\ \operatorname{cov}(\xi_2, \xi_1) & D(\xi_2) & \cdots & \operatorname{cov}(\xi_2, \xi_m) \\ \cdots & \cdots & \cdots & \cdots \\ \operatorname{cov}(\xi_m, \xi_1) & \operatorname{cov}(\xi_m, \xi_2) & \cdots & D(\xi_m) \end{pmatrix}$$

$$=\begin{pmatrix} D(\xi_1) & & & 0 \\ & D(\xi_2) & & \\ & & \ddots & \\ 0 & & & D(\xi_m) \end{pmatrix},$$

故$(\xi_1, \cdots, \xi_m)'$的概率分布密度为

$$p(x_1, \cdots, x_m)=\frac{1}{(2\pi)^{\frac{m}{2}}|B|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}\begin{pmatrix} x_1-a_1 \\ \vdots \\ x_m-a_m \end{pmatrix}' B^{-1}\begin{pmatrix} x_1-a_1 \\ \vdots \\ x_m-a_m \end{pmatrix}\right\}$$

$$=\frac{1}{\sqrt{2\pi D(\xi_1)}}e^{-\frac{(x_1-a_1)^2}{2D(\xi_1)}}\cdots\frac{1}{\sqrt{2\pi D(\xi_m)}}e^{-\frac{(x_m-a_m)^2}{2D(\xi_m)}},$$

由第二章第五节定理2(ii)知$\xi_1, \cdots, \xi_m$独立。

(ii) 也只需证明条件的充分性。

$(\xi', \eta')'$的协方差阵为

$$D\begin{pmatrix} \xi \\ \eta \end{pmatrix}=\begin{pmatrix} D(\xi) & C(\xi, \eta) \\ C(\eta, \xi) & D(\eta) \end{pmatrix}$$

$$=\begin{pmatrix} D(\xi) & 0 \\ 0 & D(\eta) \end{pmatrix}.$$

$(\xi', \eta')'$的概率分布密度为

$$p(x, y)=\frac{1}{(2\pi)^{\frac{m+n}{2}}|D(\xi)|^{\frac{1}{2}}|D(\eta)|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}\begin{pmatrix}x-E\xi\\ y-E\eta\end{pmatrix}'\begin{pmatrix}D(\xi) & 0\\ 0 & D(\eta)\end{pmatrix}^{-1}\begin{pmatrix}x-E\xi\\ y-E\eta\end{pmatrix}\right\}$$

$$=\frac{1}{(2\pi)^{\frac{m+n}{2}}|D(\xi)|^{\frac{1}{2}}|D(\eta)|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}[(x-E\xi)'(D(\xi))^{-1}(x-E\xi)\right.$$

$$\left.+(y-E\eta)'(D(\eta))^{-1}(y-E\eta)]\right\}$$

$$=\frac{1}{(2\pi)^{\frac{m}{2}}|D(\xi)|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-E\xi)'(D(\xi))^{-1}(x-E\xi)\right\}$$

$$\times\frac{1}{(2\pi)^{\frac{n}{2}}|D(\eta)|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}(y-E\eta)'(D(\eta))^{-1}(y-E\eta)\right\},$$

即可以分解为 x 的函数和 y 的函数的乘积. 由此利用第二章第五节随机向量独立的定义可以验证 ξ, η 独立.

习 题 四

1. $\xi_k\sim N(a, \sigma^2)$, $k=1, 2, \cdots, n$ 独立, 试求

$$\eta_1=\sum_{k=1}^{n}\xi_k, \qquad \eta_2=\sum_{k=1}^{m}\xi_k \quad (m<n)$$

的联合概率分布密度.

[提示: (η_1, η_2)是二维正态随机变量.]

2. 随机变量 ξ, η 有正的方差. 试证明

$$E(\xi\eta)=(E\xi)(E\eta) \quad 或 \quad D(\xi+\eta)=D\xi+D\eta$$

成立的充要条件是 ξ, η 的相关系数为 0.

3. (ξ, η)是二维正态随机变量, 有分布密度

$$p(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\right.$$
$$\left.\left[\frac{x^2}{\sigma_1^2}-2r\frac{xy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right]\right\}\quad(\sigma_1^2\neq\sigma_2^2),$$

试证明：$\xi'=\xi\cos\alpha+\eta\sin\alpha$，$\eta'=-\xi\sin\alpha+\eta\cos\alpha$ 独立，其中 α 满足

$$\operatorname{tg}2\alpha=\frac{2r\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}.$$

4. (ξ, η)在以原点为心的单位圆内均匀分布，试求 $\operatorname{cov}(\xi, \eta)$，并和第二章第五节习题 2(2)比较.

5. $(\xi, \eta)\sim N_2\left(\begin{pmatrix}a\\b\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$,
$$q=P[(\xi-a)(\eta-b)<0],$$
证明 $\rho=\cos q\pi$.

第四节　其它数字特征

为了从各个方面刻划随机变量，还有许多数字特征，如偏度，峰度，众数等等．这一节再讲两个数字特征．

4.1　中位数

定义　称满足下面条件的常数 C 为随机变量 ξ(或它的分布)的中位数：

$$P(\xi\leqslant C)\geqslant\frac{1}{2},\qquad P(\xi\geqslant C)\geqslant\frac{1}{2}.$$

【例 1】　随机变量 ξ 在[0, 1]上均匀分布，$\frac{1}{2}$ 就是 ξ 的中位数，因为

$$P\left(\xi\leqslant\frac{1}{2}\right)=\frac{1}{2},\quad P\left(\xi\geqslant\frac{1}{2}\right)=\frac{1}{2}.$$

【例 2】　随机变量 ξ 有分布列

ξ	0	1
P_ξ	$\frac{2}{3}$	$\frac{1}{3}$

0 就是 ξ 的中位数，因此

$$P(\xi\leqslant 0)=\frac{2}{3}>\frac{1}{2},\qquad P(\xi\geqslant 0)=1>\frac{1}{2}.$$

【例 3】 若随机变量 ξ 有分布列

ξ	-1	1
P_ξ	$\frac{1}{2}$	$\frac{1}{2}$

则任何满足条件 $-1<C\leqslant 1$ 的常数都是 ξ 的中位数. 因此

$$P(\xi\leqslant C)=\begin{cases}P(\xi=-1)=\dfrac{1}{2}, & \text{当 } -1<C<1;\\ P(\xi=-1 \text{ 或 } \xi=1)=1>\dfrac{1}{2}, & \text{当 } C=1.\end{cases}$$

$$P(\xi\geqslant C)=P(\xi=1)=\frac{1}{2}.$$

对任何随机变量 ξ，中位数都是存在的. 设 $F(x)$ 是 ξ 的分布函数，取

$$C=\sup\left\{x\,\middle|\,F(x)\leqslant\frac{1}{2}\right\},$$

则有

$$P(\xi\leqslant C)=F(C+0)=\lim_{0<\delta\searrow 0}F(C+\delta)\geqslant\frac{1}{2},$$

$$P(\xi\geqslant C)=1-P(\xi<C)\geqslant\frac{1}{2},$$

C 就是 ξ 的一个中位数.

当 $F(x)$ 连续时，C 是中位数的充要条件是

$$F(C)=\frac{1}{2}.$$

(证明留给读者.)

如ξ服从概率分布密度为

$$p(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$

的柯西分布,则分布函数

$$F(x)=\frac{1}{\pi}\int_{-\infty}^{x}\frac{1}{1+t^2}dt=\frac{1}{\pi}\operatorname{arc\,tg} t\ \Big|_{-\infty}^{x}$$
$$=\frac{1}{\pi}\operatorname{arc\,tg} x+\frac{1}{2}$$

连续, $F(0)=\frac{1}{2}$, 故0是中位数.

中位数是分布的"中点", 也是刻划随机变量取值大小的一个数字特征,特别对那些数学期望不存在的随机变量,这个数字特征更为重要.

4.2 熵

熵是刻划随机试验不确定性的数字特征.

考虑只包含两个基本事件A, $\overline{A}$; $P(A)=\frac{1}{2}$的随机试验,简单记作

$$\alpha=\begin{pmatrix} A & \overline{A} \\ 0.50 & 0.50 \end{pmatrix}.$$

这个随机试验是有不确定性的, 因为在一次试验中可能A出现,也可能$\overline{A}$出现. 再考虑

$$\beta=\begin{pmatrix} B & \overline{B} \\ 0.99 & 0.01 \end{pmatrix}$$

表示的随机试验,这个试验也有不确定性,由于在一次试验中

B"几乎必出现"，这个随机试验的不确定性小于前一个.

$$\gamma=\begin{pmatrix} C & \bar{C} \\ 0.7 & 0.3 \end{pmatrix}$$

表示的随机试验的不确定性则介于前两个之间. 一般的可以考虑分布列

$$P(\xi=x_i)=p_i, \quad i=1, 2, \cdots, n,$$

表示的随机试验的不确定性. 记

$$A_i=\{\xi=x_i\}, \quad i=1, 2, \cdots, n,$$

这个试验可以表示为

$$\begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}.$$

我们希望定义一个刻划这种不确定性的数值.

为了定义这个数值，我们先分析一下这个数值应当满足哪些条件.

1) 这个数仅依赖于$(p_1, \cdots, p_n)$，而且和$p_1, \cdots, p_n$的次序没有关系，因而是$p_1, \cdots, p_n$的对称函数，记为$H(p_1, \cdots, p_n)$.

2) 当$p_1, \cdots, p_n$中有一个是1时，随机试验没有不确定性，对应的$H(p_1, \cdots, p_n)$应当为0.

3) 当$p_i(1\leqslant i\leqslant n)$只有微小变化时，$H$应当也只有微小变化. 所以$H(p_1, \cdots, p_n)$对每个自变量来说应当是连续函数.

4) 当$p_1=p_2=\cdots=p_n=\dfrac{1}{n}$时，$H(p_1, \cdots, p_n)$应当有极大值.

5) $H(p_1, \cdots, p_n)$有可加性. H的可加性的意义如下：

设 $\alpha=\begin{pmatrix} A_1, & \cdots, & A_m \\ p_1, & \cdots, & p_m \end{pmatrix}$, $\beta=\begin{pmatrix} B_1, & \cdots, & B_n \\ q_1, & \cdots, & q_n \end{pmatrix}$

是两个随机试验，$p_i>0$，$i=1, 2, \cdots, n$. 考虑乘积

$$\alpha\beta=\begin{pmatrix} A_1B_1, \cdots, A_1B_n; A_2B_1, \cdots, A_2B_n; \cdots; A_mB_1, \cdots, A_mB_n \\ r_{11}, \cdots, r_{1n}; r_{21}, \cdots, r_{2n}; \cdots; r_{m1}, \cdots, r_{mn} \end{pmatrix},$$

$$r_{ij}=P(A_iB_j)=P(B_j|A_i)P(A_i)=P(B_j|A_i)p_i,$$

$$i=1, 2, \cdots, m; j=1, 2, \cdots, n.$$

把 $\alpha\beta$ 表示的随机试验看成是经过两步完成的一个试验. 第一步只看 $A_1, A_2, \cdots, A_m$ 中哪一个出现，在这一步试验有不确定性 $H(p_1, \cdots, p_m)$，但还不是整个试验的不确定性 $H(r_{11}, \cdots, r_{1n}, \cdots, r_{m1}, \cdots, r_{mn})$，因为，譬如在第一步试验中 A_i 出现，但整个试验的结果 $A_iB_1, \cdots, A_iB_n$ 究竟是哪一个出现还要看第二步试验

$$\begin{pmatrix} B_1, & B_2, & \cdots, & B_n \\ \frac{r_{i1}}{p_i}, & \frac{r_{i2}}{p_i}, & \cdots, & \frac{r_{in}}{p_i} \end{pmatrix}.$$

在第一步试验中 A_i 出现的条件下，第二步试验有不确定性

$$H\left(\frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \cdots, \frac{r_{in}}{p_i}\right),$$

我们取数学期望

$$\sum_{i=1}^{m} p_i H\left(\frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \cdots, \frac{r_{in}}{p_i}\right)$$

做为第二步试验的不确定性. 要求 H 有可加性，指的是：若试验分两步完成，整个试验的不确定性应当是这两步试验不确定性的和，也就是要求

$$H(r_{11}, \cdots, r_{1n}, \cdots, r_{m1}, \cdots, r_{mn})$$
$$=H(p_1, \cdots, p_m)+\sum_{i=1}^{m} p_i H\left(\frac{r_{i1}}{p_i}, \cdots, \frac{r_{in}}{p_i}\right).$$

以上讲的是希望刻划随机试验不确定性的函数 $H(p_1, \cdots, P_n)$ 应当满足的条件. 问题是满足这样多的条件的函数

有没有呢？回答是肯定的，例如函数

$$h(p_1, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i$$

就满足上面的条件(当 $p_i=0$ 时约定 $p_i \log p_i=0$，其中的对数为常用对数)．显然 $h(p_1, \cdots, p_n)$ 是 $p_1, \cdots, p_n$ 的对称函数、连续函数，而且某一个 $p_i=1$ 时 $h=0$．其次

$$h\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n. \qquad (1)$$

$g(x)=x\log x$ 是(下)凸函数，因而有

$$g\left(\frac{1}{n} \sum_{i=1}^{n} p_i\right) \leqslant \frac{1}{n} \sum_{i=1}^{n} g(p_i).$$

注意到 $\sum\limits_{i=1}^{n} p_i=1$，上式即

$$\frac{1}{n} \log \frac{1}{n} \leqslant \frac{1}{n} \sum_{i=1}^{n} p_i \log p_i,$$

$$\log n \geqslant -\sum_{i=1}^{n} p_i \log p_i.$$

和(1)式比较得

$$h\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) \geqslant -\sum_{i=1}^{n} p_i \log p_i.$$

最后，若将试验

$$\alpha\beta = \begin{pmatrix} A_1B_1, & \cdots, & A_1B_n; & \cdots; & A_mB_1, & \cdots, & A_mB_n \\ r_{11}, & \cdots, & r_{1n}; & \cdots; & r_{m1}, & \cdots, & r_{mn} \end{pmatrix}$$

分为两步，第一步试验为

$$\alpha = \begin{pmatrix} A_1, & A_2, & \cdots, & A_m \\ p_1, & p_2, & \cdots, & p_m \end{pmatrix},$$

$$p_i - \sum_{j=1}^{n} r_{ij} > 0, \quad i=1, 2, \cdots, m,$$

则 $h(r_{11}, \cdots, r_{1n}; \cdots, r_{m1}, \cdots, r_{mn}) = -\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}\log r_{ij}$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}\left(\log p_i + \log\frac{r_{ij}}{p_i}\right)$$

$$= -\sum_{i=1}^{m}\left(\sum_{j=1}^{n} r_{ij}\right)\log p_i - \sum_{i=1}^{m} p_i \sum_{j=1}^{n}\frac{r_{ij}}{p_i}\log\frac{r_{ij}}{p_i}$$

$$= -\sum_{i=1}^{m} p_i\log p_i + \sum_{i=1}^{m} p_i h\left(\frac{r_{i1}}{p_i}, \cdots, \frac{r_{in}}{p_i}\right)$$

$$= h(p_1, \cdots, p_m) + \sum_{i=1}^{m} p_i h\left(\frac{r_{i1}}{p_i}, \cdots, \frac{r_{in}}{p_i}\right).$$

定义 在具有离散分布

$$\alpha = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

的随机试验中，称

$$H(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i\log p_i$$

为 α 的熵.

其中 log 为常用对数. 又若某一 p_i 为 0, 约定 $p_i\log p_i$ 为 0.

例如，

$$H(0.99, 0.01) = -(0.99\times\log 0.99 + 0.01\times\log 0.01) = 0.0243,$$

$$H(0.70, 0.30) = -(0.70\times\log 0.70 + 0.30\times\log 0.30) = 0.2653,$$

$$H(0.50, 0.50) = 0.3010.$$

以上我们对只取有限个值的离散分布定义了熵，并讨论了熵的性质. 可以证明：满足条件 1～5 的 H 的形式是唯一的，并且可以在更弱的条件下被确定. 对连续分布也可以定义熵.

第五节　特征函数

特征函数是研究随机变量分布律的一个重要工具．有时直接确定一个随机变量的分布律比较困难，但确定它的特征函数比较容易，由于分布律和特征函数之间有一一对应的关系，在得知特征函数后就可以知道分布律．

在这一节将用到取复数值的随机变量．

如果 ξ, η 是二维实随机变量，就称 $\xi+i\eta$ 为复随机变量，其中 $i^2=-1$.

如果二维实随机变量 $(\xi_1,\ \eta_1)$, $(\xi_2,\ \eta_2)$, $\cdots$, $(\xi_n,\ \eta_n)$ 独立，就称复随机变量 $\xi_1+i\eta_1$, $\xi_2+i\eta_2$, $\cdots$, $\xi_n+i\eta_n$ 独立．

如果二维实随机变量 ξ, η 的数学期望 $E\xi$, $E\eta$ 存在，就称数 $E\xi+iE\eta$ 为复随机变量 $\xi+i\eta$ 的数学期望，记作 $E(\xi+i\eta)$．复随机变量的数学期望也有第三章1.2中讲的几个性质，读者不妨自己证明一下．这里对复随机变量证明

性质 11　　$|E(\xi+i\eta)|\leqslant E|\xi+i\eta|$

【证】 设　$E(\xi+i\eta)=|E(\xi+i\eta)|e^{i\varphi}$,

$$\xi+i\eta=|\xi+i\eta|e^{i\phi(\xi,\eta)},$$

则

$$\begin{aligned}|E(\xi+i\eta)|&=e^{-i\varphi}E(\xi+i\eta)\\&=e^{-i\varphi}E[|\xi+i\eta|e^{i\phi(\xi,\eta)}]\\&=E\{|\xi+i\eta|e^{i[\phi(\xi,\eta)-\varphi]}\}\\&=E\{|\xi+i\eta|\cos[\phi(\xi,\ \eta)-\varphi]\}\\&\quad+iE\{|\xi+i\eta|\sin[\phi(\xi,\ \eta)-\varphi]\}.\end{aligned}$$

上式左端是实数，故右端虚部为零，由第一节1.2推论3

$$|E(\xi+i\eta)|\leqslant E|\xi+i\eta|.\quad ▌$$

5.1 特征函数的定义及初等性质

定义 ξ 是随机变量，称复随机变量 $e^{it\xi}=\cos t\xi+i\sin t\xi$ 的数学期望

$$f_\xi(t)=Ee^{it\xi}=E\cos t\xi+iE\sin t\xi \tag{1}$$

为 ξ 的特征函数，其中 t 是实数。

因为对任何实数 t, $E|\cos t\xi|\leqslant 1$, $E|\sin t\xi|\leqslant 1$, 所以 $f_\xi(t)$ 对一切实数有定义。

若 ξ 有分布列

$$P(\xi=x_k)=p_k,\quad k=1,2,\cdots.$$

$$f_\xi(t)=\sum_k(\cos tx_k)p_k+i\sum_k(\sin tx_k)p_k=\sum_k e^{itx_k}p_k. \tag{2}$$

若 ξ 有概率分布密度 $p(x)$,

$$f_\xi(t)=\int_{-\infty}^{+\infty}(\cos tx)p(x)dx+i\int_{-\infty}^{+\infty}(\sin tx)p(x)dx$$

$$=\int_{-\infty}^{+\infty}e^{itx}p(x)dx. \tag{3}$$

【例 1】 ξ 有分布列 $\begin{pmatrix}0 & 1\\ 1-p & p\end{pmatrix}$，试求 ξ 的特征函数 $f_\xi(t)$.

解：由(2)

$$f_\xi(t)=(1-p)+e^{it}p.$$

【例 2】 ξ 服从参数为 λ 的泊松分布，试求 ξ 的特征函数 $f_\xi(t)$.

解：ξ 有分布列

$$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots,$$

由(2)

$$f_\xi(t)=\sum_{k=0}^{\infty}e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!}$$

$$=e^{-\lambda}e^{\lambda e^{it}}=e^{\lambda(e^{it}-1)}.$$

【例 3】 $\xi\sim N(0,1)$，试求 ξ 的特征函数.

解：由(3)

$$f_\xi(t)=\int_{-\infty}^{+\infty}\cos tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$+i\int_{-\infty}^{+\infty}\sin tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,$$

等式右端第二项为 0，将 $\cos tx$ 展成级数并逐项积分*)

$$f_\xi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sum_{k=0}^{\infty}\frac{(tx)^{2k}}{(2k)!}(-1)^k e^{-\frac{x^2}{2}}dx$$

$$=\frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty}\frac{(-t^2)^k}{(2k)!}\int_{-\infty}^{+\infty}x^{2k}e^{-\frac{x^2}{2}}dx,$$

由第一节例 6 知

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^{2k}e^{-\frac{x^2}{2}}dx=(2k-1)(2k-3)\cdots5\cdot3\cdot1.$$

---

*)
$$\left|\int_{-\infty}^{+\infty}(\cos tx)e^{-\frac{x^2}{2}}dx-\sum_{k=0}^{N}\int_{-\infty}^{+\infty}\frac{(tx)^{2k}}{(2k)!}(-1)^k e^{-\frac{x^2}{2}}dx\right|$$

$$\leqslant 2\int_0^{+\infty}\left|\cos tx-\sum_{k=0}^{N}\frac{(tx)^{2k}}{(2k)!}(-1)^k\right|e^{-\frac{x^2}{2}}dx$$

$$\leqslant 2\int_0^{A}\left|\cos tx-\sum_{k=0}^{N}\frac{(tx)^{2k}}{(2k)!}(-1)^k\right|e^{-\frac{x^2}{2}}dx$$

$$+2\int_A^{+\infty}|\cos tx|e^{-\frac{x^2}{2}}dx+2\int_A^{+\infty}\frac{1}{2}(e^{tx}+e^{-tx})e^{-\frac{x^2}{2}}dx,$$

任给定 $\varepsilon>0$，取充分大的 A 使后两个积分都不大于 $\frac{\varepsilon}{3}$. 第一个积分的被积函数在$[0, A]$上当 $N\to+\infty$ 时一致收敛于0，所以当 N 充分大时，第一个积分不大于 $\frac{\varepsilon}{3}$.

所以 $f_\xi(t)=\sum_{k=0}^{+\infty}\frac{(-t^2)^k}{(2k)!}(2k-1)(2k-3)\cdots5\cdot3\cdot1$

$$=\sum_{k=0}^{+\infty}\frac{(-t^2)^k}{2^k k!}=\sum_{k=0}^{\infty}\frac{\left(-\frac{t^2}{2}\right)^k}{k!}=e^{-\frac{t^2}{2}}.$$

现在,我们导出特征函数的初等性质.

如果没有特别说明,以下的 $f_\xi(t)$ 都表示随机变量 ξ 的特征函数.

性质 1 $f_\xi(t)$ 有下列简单性质

(i) $f_\xi(0)=1$;

(ii) 对任何实数 t, $|f_\xi(t)|\leqslant1$;

(iii) $f_\xi(-t)=\overline{f_\xi(t)}$.

【证】 (i) 是显然的. 对任何实数 t, 由性质 11,

$$|f_\xi(t)|=|Ee^{it\xi}|\leqslant E|e^{it\xi}|=1,$$

所以(ii)成立. 由(1)

$$\begin{aligned}f_\xi(-t)&=E\cos(-t)\xi+iE\sin(-t)\xi\\&=E\cos t\xi-iE\sin t\xi\\&=\overline{E\cos t\xi+iE\sin t\xi}=\overline{f_\xi(t)},\end{aligned}$$

所以(iii)成立. **】**

性质 2 a, b 是常数,则 $f_{a\xi+b}(t)=e^{itb}f_\xi(at)$.

【证】 $$\begin{aligned}f_{a\xi+b}(t)&=Ee^{it(a\xi+b)}=Ee^{i(at)\xi}\cdot e^{itb}\\&=e^{itb}\cdot Ee^{i(at)\xi}=e^{itb}f_\xi(at).\end{aligned}$$ **】**

性质 3 $f_\xi(t)$ 在实直线上一致连续.

【证】 对任何实数 t 及 h,

$$\begin{aligned}|f_\xi(t+h)-f_\xi(t)|&=|Ee^{i(t+h)\xi}-Ee^{it\xi}|\\&=\left|Ee^{i\left(t+\frac{h}{2}\right)\xi}\left(e^{i\frac{h\xi}{2}}-e^{-i\frac{h\xi}{2}}\right)\right|\\&\leqslant E\left|e^{i\frac{h\xi}{2}}-e^{-i\frac{h\xi}{2}}\right|=2E\left|\sin\frac{h\xi}{2}\right|,\end{aligned}$$

令
$$\eta=\begin{cases}1, & 当|\xi|>A,\\ \dfrac{|h|A}{2}, & 当|\xi|\leqslant A,\end{cases}$$

则
$$\left|\sin\frac{h\xi}{2}\right|\leqslant\eta,$$

因而
$$\begin{aligned}2E\left|\sin\frac{h\xi}{2}\right|&\leqslant 2E\eta\\ &=|h|AP(|\xi|\leqslant A)+2P(|\xi|>A)\\ &\leqslant|h|A+2P(|\xi|>A).\end{aligned}$$

任给定 $\varepsilon>0$, 选一个充分大的 A 使
$$2P(|\xi|>A)<\frac{\varepsilon}{2}.$$

当 $|h|\leqslant\dfrac{\varepsilon}{2A}$ 时,
$$|f_\xi(t+h)-f_\xi(t)|<\varepsilon$$
对 t 一致成立, 所以 $f_\xi(t)$ 一致连续. ▌

性质 4 对任何自然数 n, 实数 $t_1, t_2, \cdots, t_n$, 及复数 $u_1, u_2, \cdots, u_n$,
$$\sum_{j=1}^{n}\sum_{k=1}^{n}f_\xi(t_j-t_k)u_j\bar{u}_k\geqslant 0.$$

【证】
$$\begin{aligned}\sum_{j=1}^{n}\sum_{k=1}^{n}f_\xi(t_j-t_k)u_j\bar{u}_k&=\sum_{j=1}^{n}\sum_{k=1}^{n}Ee^{i(t_j-t_k)\xi}u_j\bar{u}_k\\ &=E\left[\sum_{j=1}^{n}\sum_{k=1}^{n}(u_je^{it_j\xi})(\overline{u_ke^{it_k\xi}})\right]=E\left|\sum_{j=1}^{n}u_je^{it_j\xi}\right|^2\geqslant 0.\end{aligned}$$ ▌

称具有性质 4 的函数 $f(t)$ 为在实直线上正定.

连续, 在原点的值为 1, 在实直线上正定是特征函数的必要条件, 也是充分条件. 后者的证明在《测度论与概率论基础》中可以找到.

性质 5 若ξ, η独立,则$f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$.

【证】 ξ, η独立时,复随机变量$e^{it\xi}$, $e^{it\eta}$独立.

$$f_{\xi+\eta}(t)=Ee^{it(\xi+\eta)}=Ee^{it\xi}\cdot e^{it\eta}$$
$$=(Ee^{it\xi})(Ee^{it\eta})=f_\xi(t)\cdot f_\eta(t). \quad \rrbracket$$

这个性质可以推广为：若ξ_1, ξ_2, $\cdots$, ξ_n独立,则

$$f_{\sum\limits_{k=1}^{n}\xi_k}(t)=\prod_{k=1}^{n}f_{\xi_k}(t).$$

性质 6 随机变量ξ有n阶绝对矩，则ξ的特征函数可以微分n次,并且在$k\leqslant n$时

$$f_\xi^{(k)}(0)=i^kE\xi^k.$$

以下就ξ是连续型随机变量的情况给以证明. 离散型情况的证明是类似的. 这个性质对一般随机变量也是对的.

设ξ有概率分布密度$p(x)$. 因为ξ有n阶绝对矩，积分

$$\int_{-\infty}^{+\infty}|x|^kp(x)dx, \quad k=1, 2, \cdots, n \tag{4}$$

收敛. 令

$$f_1(t)=\int_{-\infty}^{+\infty}(\cos tx)p(x)dx,$$

$$f_2(t)=\int_{-\infty}^{+\infty}(\sin tx)p(x)dx,$$

则
$$f_\xi(t)=f_1(t)+if_2(t),$$

$$\frac{f_1(t+\Delta t)-f_1(t)}{\Delta t}$$
$$=\int_{-\infty}^{+\infty}\frac{\cos(t+\Delta t)x-\cos tx}{\Delta t}p(x)dx$$
$$=-\int_{-\infty}^{+\infty}x\sin([t+\theta(x)\Delta t]x)\cdot p(x)dx,$$
$$|\theta(x)|\leqslant 1. \tag{5}$$

现在证明：当$\Delta t\to 0$时,(5)式以

$$-\int_{-\infty}^{+\infty} x(\sin tx)p(x)dx$$

为极限.

$$\left|\frac{f_1(t+\Delta t)-f_1(t)}{\Delta t}+\int_{-\infty}^{+\infty} xp(x)\sin tx\,dx\right|$$

$$=\left|\int_{-\infty}^{+\infty} xp(x)\sin tx\,dx-\int_{-\infty}^{+\infty} xp(x)\sin([t+\theta(x)\Delta t]x)dx\right|$$

$$\leqslant\int_{-A}^{A}|x|p(x)\,|\sin tx-\sin[t+\theta(x)\Delta t]x|\,dx$$

$$+2\int_{|x|>A}|x|p(x)dx \tag{6}$$

任给定 $\varepsilon>0$, 由(4), 取充分大的 A 可以使

$$2\int_{|x|>A}|x|p(x)dx<\frac{\varepsilon}{2}.$$

对$[-A, A]$上的 x,

$$|\sin tx-\sin[t+\theta(x)\Delta t]x|$$

$$=2\left|\cos\left(\frac{2t+\theta(x)\Delta t}{2}x\right)\sin\frac{\theta(x)\Delta t}{2}x\right|$$

$$\leqslant|\Delta t|\,|x|=A|\Delta t|.$$

当$|\Delta t|<\dfrac{\varepsilon}{2A^2}$时,

$$\int_{-A}^{A}|x|p(x)\,|\sin tx-\sin[t+\theta(x)\Delta t]x|\,dx$$

$$\leqslant A^2|\Delta t|<\frac{\varepsilon}{2}.$$

所以 $f_1'(t)$存在,并且

$$f_1'(t)=-\int_{-\infty}^{+\infty} x(\sin tx)p(x)dx,$$

即可以在积分号内微分.

用同样方法可以证明 $f_2'(t)$存在, 也可以在积分号内微分,

$$f_2'(t)=\int_{-\infty}^{+\infty}x(\cos tx)p(x)dx.$$

所以

$$f'(t)=f_1'(t)+if_2'(t)=i[f_2'(t)-if_1'(t)]$$
$$=i\Big[\int_{-\infty}^{+\infty}x(\cos tx)p(x)dx$$
$$+i\int_{-\infty}^{+\infty}x(\sin tx)p(x)dx\Big], \tag{7}$$
$$f'(0)=iE\xi.$$

对 $f'(t)$ 重复上面的步骤可以证明 $f''(t)$ 存在，并且可以在(8)式积分号内微分，

$$f''(t)=i^2\Big[\int_{-\infty}^{+\infty}x^2(\cos tx)p(x)dx$$
$$+i\int_{-\infty}^{+\infty}x^2(\sin tx)p(x)dx\Big],$$
$$f''(0)=i^2E\xi^2.$$

再对 $f''(t)$ 重复上面的步骤……．这样，性质6中 ξ 为连续型随机变量的情况得到证明．

【例 4】 $\xi\sim N(a,\ \sigma^2)$，试求 $f_\xi(t)$．

解： 设 $\eta=\dfrac{\xi-a}{\sigma}$，则 $\eta\sim N(0,\ 1)$，并且 $\xi=\sigma\eta+a$．由性质2及例3

$$f_\xi(t)=f_{\sigma\eta+a}(t)=e^{ita}f_\eta(\sigma t)=e^{ita}e^{-\frac{(\sigma t)^2}{2}}=e^{iat-\frac{\sigma^2t^2}{2}}.$$

【例 5】 $\xi\sim N(0,\ 1)$，$\eta=\xi^2$，试求 η 的特征函数 $f_\eta(t)$．

解： η 是自由度为1的 χ^2 变量，有概率分布密度

$$p(x)=\begin{cases}0, & x\leqslant 0;\\ \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}x^{-\frac{1}{2}}, & x>0.\end{cases}$$

由特征函数的定义

$$
\begin{aligned}
f_\eta(t) &= E\cos t\eta + iE\sin t\eta \\
&= \int_0^{+\infty} \cos tx \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}} dx \\
&\quad + i\int_0^{+\infty} \sin tx \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}} dx \\
&= f_1(t) + if_2(t).
\end{aligned}
$$

从便于计算考虑，先计算 $f_\eta^2(t)$.

$$f_\eta^2(t) = f_1^2(t) - f_2^2(t) + 2if_1(t)f_2(t),$$

$$
\begin{aligned}
f_1^2(t) &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \cos tx e^{-\frac{x}{2}} x^{-\frac{1}{2}} dx \\
&\quad \times \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \cos ty e^{-\frac{y}{2}} y^{-\frac{1}{2}} dy \\
&= \frac{1}{2\pi} \int_0^{+\infty}\int_0^{+\infty} \cos tx \cos ty e^{-\frac{x+y}{2}} (xy)^{-\frac{1}{2}} dx dy,
\end{aligned}
$$

$$f_2^2(t) = \frac{1}{2\pi} \int_0^{+\infty}\int_0^{+\infty} \sin tx \sin ty e^{-\frac{x+y}{2}} (xy)^{-\frac{1}{2}} dx dy,$$

$$
\begin{aligned}
& f_1^2(t) - f_2^2(t) \\
&= \frac{1}{2\pi} \int_0^{+\infty}\int_0^{+\infty} \cos t(x+y) e^{-\frac{x+y}{2}} (xy)^{-\frac{1}{2}} dx dy. \qquad (8)
\end{aligned}
$$

令 $\quad x = \rho\cos^2\theta, \quad y = \rho\sin^2\theta \quad \left(\rho > 0,\ 0 \leqslant \theta \leqslant \frac{\pi}{2}\right),$

则 $\quad \left|\frac{\partial(x, y)}{\partial(\rho, \theta)}\right| = 2\rho\sin\theta\cos\theta = 2(xy)^{\frac{1}{2}}.$

代入(8)式

$$
\begin{aligned}
f_1^2(t) - f_2^2(t) &= \frac{1}{2\pi} \int_0^{+\infty}\int_0^{\frac{\pi}{2}} \cos t\rho e^{-\frac{\rho}{2}} \cdot 2\, d\theta\, d\rho \\
&= \frac{1}{2} \int_0^{+\infty} \cos t\rho e^{-\frac{\rho}{2}} d\rho,
\end{aligned}
$$

用分部积分法不难算出

$$\int_0^{+\infty} \cos t\rho e^{-\frac{\rho}{2}} d\rho = \frac{2}{1+4t^2}.$$

所以

$$f_1^2(t)-f_2^2(t)=\frac{1}{1+4t^2}.$$

类似的有

$$\begin{aligned}
&2f_1(t)f_2(t)\\
&=\frac{1}{2\pi}\int_0^{+\infty}\int_0^{+\infty}(\cos tx)(\sin ty)e^{-\frac{x+y}{2}}(xy)^{-\frac{1}{2}}dx\,dy\\
&+\frac{1}{2\pi}\int_0^{+\infty}\int_0^{+\infty}(\sin tx)(\cos ty)e^{-\frac{x+y}{2}}(xy)^{-\frac{1}{2}}dx\,dy\\
&=\frac{1}{2\pi}\int_0^{+\infty}\int_0^{+\infty}\sin t(x+y)\,e^{-\frac{x+y}{2}}(xy)^{-\frac{1}{2}}dx\,dy\\
&=\frac{1}{2}\int_0^{+\infty}\sin t\rho\, e^{-\frac{\rho}{2}}d\rho=\frac{2t}{1+4t^2},
\end{aligned}$$

所以

$$f_\eta^2(t)=\frac{1}{1+4t^2}+i\,\frac{2t}{1+4t^2}=\frac{1+2it}{1+4t^2}=\frac{1}{1-2it},$$

$$f_\eta(t)=(1-2it)^{-\frac{1}{2}}.$$

注意, $f_\eta(0)=1$, $f_\eta(t)$连续.

【例 6】 ξ是自由度为n的χ^2变量,试求$f_\xi(t)$.

解: 设$\xi_1, \xi_2, \cdots, \xi_n$是独立标准正态随机变量,则

$$\eta=\xi_1^2+\xi_2^2+\cdots+\xi_n^2$$

是自由度为n的χ^2变量,与ξ有相同的分布. 由性质5及例题5

$$\begin{aligned}
f_\xi(t)&=f_\eta(t)=\prod_{k=1}^n f_{\xi_k^2}(t)\\
&=\prod_{k=1}^n(1-2it)^{-\frac{1}{2}}=(1-2it)^{-\frac{n}{2}}.
\end{aligned}$$

【例 7】 $\xi'=(\xi_1, \xi_2, \cdots, \xi_n)$是由独立标准正态随机变量为分量组成的向量, $A=(a_{ij})$是$n\times n$实对称方阵,已知二

次型 $\xi' A\xi$ 是自由度为 m 的 χ^2 变量，试证 A 是幂等矩阵(即 $A^2=A$)并且秩 $A=m$.

【证】 对实对称矩阵 A 来说，有一个正交方阵 C 使

$$C'AC=\begin{bmatrix} \lambda_1 & & & & & 0 \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{bmatrix},$$

$\lambda_1, \lambda_2, \cdots, \lambda_r$ 不为 0. 设 $\xi' C=(\eta_1, \eta_2, \cdots, \eta_n)=\eta'$，则 η_1, $\eta_2, \cdots, \eta_n$ 仍是独立的标准正态随机变量.

$$\begin{aligned} \xi' A\xi &= \xi' CC' ACC'\xi=\eta'(C'AC)\eta \\ &= \lambda_1\eta_1^2+\lambda_2\eta_2^2+\cdots+\lambda_r\eta_r^2. \end{aligned} \tag{9}$$

已知 $\xi' A\xi$ 是自由度为 m 的 χ^2 变量，由例 6，有特征函数 $(1-2it)^{-\frac{m}{2}}$. 由(9)，$\xi' A\xi$ 的特征函数是

$$f_{\sum\limits_{k=1}^{r}\lambda_k\eta_k^2}(t)=\prod_{k=1}^{r} f_{\lambda_k\eta_k^2}(t)=\prod_{k=1}^{r}(1-2i\lambda_k t)^{-\frac{1}{2}},$$

故

$$\prod_{k=1}^{r}(1-2i\lambda_k t)^{-\frac{1}{2}}=(1-2it)^{-\frac{m}{2}},$$

$$\prod_{k=1}^{r}(1-2i\lambda_k t)=(1-2it)^m.$$

对任何实数 t, $2it$ 都是多项式

$$\prod_{k=1}^{r}(1-\lambda_k z)-(1-z)^m \tag{10}$$

的零点，故(10)是零多项式，于是 $r=m$. 由于 $\frac{1}{\lambda_k}(1\leqslant k\leqslant r)$ 是(10)的零点，因而

$$\left(1-\frac{1}{\lambda_k}\right)^m=0, \quad \lambda_k=1,$$

$$A^2=C\begin{bmatrix}\lambda_1 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{bmatrix}C'C$$

$$\times\begin{bmatrix}\lambda_1 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{bmatrix}C'$$

$$=C\begin{bmatrix}\lambda_1 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{bmatrix}C'=A,$$

秩 $A=r=m$. ∎

5.2 逆转公式及唯一性定理

从特征函数的定义已经看到，随机变量的分布律（离散型随机变量的分布列、连续型随机变量的分布密度）唯一决定随机变量的特征函数. 这一小节将证明随机变量的特征函数也唯一决定随机变量的分布律.

定理1(唯一性定理) 若随机变量 ξ 的特征函数为 $f(t)$, 则 ξ 的分布函数 $F(x)$ 在其连续点 x_0 上的值

$$F(x_0)=\frac{1}{2\pi}\lim_{y\to-\infty}\lim_{T\to+\infty}\int_{-T}^{T}\frac{e^{-ity}-e^{-itx_0}}{it}f(t)dt.$$

这个定理说明 ξ 的分布函数 $F(x)$ 在连续点上的值由 $f(t)$ 决定. 由于 $F(x)$ 是单调左连续的, $F(x)$ 在不连续点上的值由连续点上的值决定, 所以 $F(x)$ 完全由 $f(t)$ 所决定. 定理1是下面定理2的直接结果.

定理2(逆转公式) 随机变量 ξ 的特征函数是 $f(t)$, $x_1<x_2$ 是 ξ 的分布函数 $F(x)$ 的两个连续点, 则

$$F(x_2)-F(x_1)=\frac{1}{2\pi}\lim_{T\to+\infty}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)dt. \quad (11)$$

【证】 1° 我们就 ξ 是连续型随机变量的情况证明(11), 离散型情况类似.

设 ξ 有概率分布密度 $p(x)$, 则对任何 $T>0$,

$$\begin{aligned}J_T&=\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)dt\\&=\int_{-T}^{T}dt\int_{-\infty}^{\infty}\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}p(x)dx.\end{aligned}$$

现在证明上述累次积分可以调换顺序. 令

$$g(t,\ x;x_1,\ x_2)=\frac{1}{it}\left(e^{it(x-x_1)}-e^{it(x-x_2)}\right),$$

则 $g(t,\ x;x_1,\ x_2)$ 是 $(t,\ x)$ 的连续函数且

$$\begin{aligned}&|g(t,\ x;x_1,\ x_2)|\\&=\left|e^{it[x-(x_1+x_2)/2]}\right|\left|\frac{1}{it}\left[e^{it(x_2-x_1)/2}-e^{it(x_1-x_2)/2}\right]\right|\\&=\left|\frac{2}{t}\sin\frac{t(x_2-x_1)}{2}\right|\\&\leqslant|x_2-x_1|,\end{aligned}$$

$$\left|J_T-\int_{-A}^{A}\left[\int_{-T}^{T}g(t,\ x;x_1,\ x_2)dt\right]p(x)dx\right|$$

$$=\left|\int_{-T}^{T}\left[\int_{-\infty}^{\infty}g(t,\ x;x_1,\ x_2)p(x)dx\right]dt\right.$$

$$\left.-\int_{-T}^{T}\left[\int_{-A}^{A}g(t,\ x;x_1,\ x_2)p(x)dx\right]dt\right|$$

$$\leqslant\int_{-T}^{T}\left[\int_{|x|>A}|g(t,\ x;x_1,\ x_2)|p(x)dx\right]dt$$

$$\leqslant 2|x_2-x_1|T\int_{|x|>A}p(x)dx.$$

由于

$$\lim_{A\to\infty}\int_{|x|>A}p(x)dx=0,$$

在上述不等式中令 $A\to\infty$，则右端趋近于 0，从而有

$$J_T=\int_{-\infty}^{\infty}\left[\int_{-T}^{T}\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}dt\right]p(x)dx.$$

再由 $\frac{1}{t}\cos t\alpha$ 是奇函数，$\frac{1}{t}\sin t\alpha$ 是偶函数知

$$J_T=2\int_{-\infty}^{\infty}\left[\int_{0}^{T}\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]p(x)dx. \tag{12}$$

2° 要想证明定理，由 (11) 知就是要对充分大的 T，估计

$$J_T-2\pi[F(x_2)-F(x_1)]$$

$$=J_T-2\pi\int_{x_1}^{x_2}p(x)dx. \tag{13}$$

为此将(12)中 J_T 对 x 的积分区间$(-\infty,\ \infty)$分成五段$(-\infty,\ x_1-\delta)$，$(x_1-\delta,\ x_1+\delta)$，$(x_1+\delta,\ x_2-\delta)$，$(x_2-\delta,\ x_2+\delta)$，$(x_2+\delta,\ +\infty)$，其中 $\delta>0$ 满足 $x_1+\delta<x_2-\delta$，并引用数学分

析中的下列结果：

(i) 对任何 $\delta>0$ 及 $\varepsilon>0$, 存在 $T_0>0$, 使当 $T>T_0$ 时,

$$\left|\int_0^T \frac{\sin \alpha t}{t}dt-\frac{\pi}{2}\right|<\varepsilon, \quad 对一切 \alpha\geqslant\delta; \tag{14}$$

且

$$\left|\int_0^T \frac{\sin \alpha t}{t}dt+\frac{\pi}{2}\right|<\varepsilon, \quad 对一切 \alpha\leqslant-\delta \tag{15}$$

成立. (ii)对一切实数 α, T

$$\left|\int_0^T \frac{\sin \alpha t}{t}dt\right|\leqslant\pi. \tag{16}$$

(这些结果的证明将在本定理证明后给出.)

任给 $\varepsilon>0$, 取 $\delta>0$ 使 $x_1+\delta<x_2-\delta$, 并且满足

$$\max_{i=1,2}\int_{x_i-\delta}^{x_i+\delta} p(x)dx<\varepsilon. \tag{17}$$

再取满足(i)的 T_0. 于是当 $x\leqslant x_1-\delta$ 时, $x-x_1\leqslant-\delta$, 从而当 $T>T_0$ 时, 由(15)知

$$\begin{aligned}&\left|\int_{-\infty}^{x_1-\delta}\left[\int_0^T\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]p(x)dx\right|\\ &\leqslant\int_{-\infty}^{x_1-\delta}\left[\left|\int_0^T\frac{\sin t(x-x_1)}{t}dt+\frac{\pi}{2}\right|\right.\\ &\quad\left.+\left|\int_0^T\frac{\sin t(x-x_2)}{t}dt+\frac{\pi}{2}\right|\right]p(x)dx\\ &\leqslant 2\varepsilon\int_{-\infty}^{x_1-\delta}p(x)dx\leqslant 2\varepsilon.\end{aligned} \tag{18}$$

同样由(14)知当 $T>T_0$ 时,

$$\left|\int_{x_2+\delta}^{\infty}\left[\int_0^T\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]p(x)dx\right| \leqslant 2\varepsilon. \tag{19}$$

由(16), (17)即得

$$\left|\int_{x_i-\delta}^{x_i+\delta}\left[\int_0^T\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]p(x)dx\right|$$
$$\leqslant\int_{x_i-\delta}^{x_i+\delta}\left[\left|\int_0^T\frac{\sin t(x-x_1)}{t}dt\right|\right.$$
$$\left.+\left|\int_0^T\frac{\sin t(x-x_2)}{t}dt\right|\right]p(x)dx$$
$$\leqslant 2\pi\int_{x_i-\delta}^{x_i+\delta}p(x)dx<2\pi\varepsilon,\quad i=1,\ 2. \tag{20}$$

最后，当 $x_1+\delta\leqslant x\leqslant x_2-\delta$ 时，$x-x_1\geqslant\delta$, $x-x_2\leqslant-\delta$，于是由(14)，(15)及(17)得

$$\left|\int_{x_1+\delta}^{x_2-\delta}\left[\int_0^T\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]p(x)dx\right.$$
$$\left.-\pi\int_{x_1}^{x_2}p(x)dx\right|$$
$$\leqslant\int_{x_1+\delta}^{x_2-\delta}\left[\left|\int_0^T\frac{\sin t(x-x_1)}{t}dt-\frac{\pi}{2}\right|\right.$$
$$\left.+\left|\int_0^T\frac{\sin t(x-x_2)}{t}dt+\frac{\pi}{2}\right|\right]p(x)dx$$
$$+\pi\int_{x_1}^{x_1+\delta}p(x)dx+\pi\int_{x_2-\delta}^{x_2}p(x)dx$$
$$\leqslant 2\varepsilon\int_{x_1+\delta}^{x_2-\delta}p(x)dx+\pi\varepsilon+\pi\varepsilon\leqslant 2(\pi+1)\varepsilon. \tag{21}$$

综合(13)，(12)及(18)～(21)即得：对任给 $\varepsilon>0$，存在 T_0，使当 $T>T_0$ 时，

$$|J_T-2\pi[F(x_2)-F(x_1)]|$$
$$<2\varepsilon+2\varepsilon+2\times 2\pi\varepsilon+2(\pi+1)\varepsilon=6(\pi+1)\varepsilon.$$

故(11)成立． ▮

注 现在我们给出(14)～(16)的证明．由数学分析知

$$\int_0^{+\infty}\frac{\sin t}{t}dt=\frac{\pi}{2},$$

即对任何 $\varepsilon>0$, 存在 $T_1>0$, 当 $T>T_1$ 时,

$$\left|\int_0^T \frac{\sin t}{t}dt-\frac{\pi}{2}\right|<\varepsilon.$$

取 $T_0=\frac{T_1}{\delta}$, 则当 $T>T_0$ 时, 对一切 $\alpha\geqslant\delta$, $\alpha T>\delta\cdot\frac{T_1}{\delta}=T_1$, 从而上式即得

$$\left|\int_0^T \frac{\sin\alpha t}{t}dt-\frac{\pi}{2}\right|=\left|\int_0^{\alpha T}\frac{\sin t}{t}dt-\frac{\pi}{2}\right|<\varepsilon,$$

此即(14); 同样可证(15).

对于任何实数 α, T, 存在一非负整数 k, 使

$$k\pi\leqslant|\alpha T|<(k+1)\pi,$$

于是

$$\int_0^{|\alpha T|}\frac{\sin t}{t}dt=\sum_{r=0}^{k-1}\int_{r\pi}^{(r+1)\pi}\frac{\sin t}{t}dt+\int_{k\pi}^{|\alpha T|}\frac{\sin t}{t}dt$$

$$=\sum_{r=0}^{k-1}(-1)^r\int_0^\pi\frac{\sin t}{r\pi+t}dt$$

$$+(-1)^k\int_0^{|\alpha T|-k\pi}\frac{\sin t}{k\pi+t}dt.$$

因为

$$\int_0^\pi\frac{\sin t}{r\pi+t}dt\geqslant\int_0^\pi\frac{\sin t}{(r+1)\pi+t}dt$$

$$(r=0,\ 1,\ \cdots,\ k-2),$$

$$\int_0^{|\alpha T|-k\pi}\frac{\sin t}{k\pi+t}dt\leqslant\int_0^\pi\frac{\sin t}{k\pi+t}dt.$$

所以

$$0\leqslant\int_0^{|\alpha T|}\frac{\sin t}{t}dt\leqslant\int_0^\pi\frac{\sin t}{t}dt\leqslant\pi,$$

$$\therefore\ \left|\int_0^T\frac{\sin\alpha t}{t}dt\right|=\left|\int_0^{\alpha T}\frac{\sin t}{t}dt\right|=\int_0^{|\alpha T|}\frac{\sin t}{t}dt\leqslant\pi,$$

故(16)获证.

【例 8】 若 $\xi_k \sim N(a_k, \sigma_k^2)$，$k=1, 2, \cdots, n$ 独立，则 $\sum_{k=1}^{n} \xi_k \sim N\left(\sum_{k=1}^{n} a_k, \sum_{k=1}^{n} \sigma_k^2\right)$.

【证】 由性质 5 及例 4，$\sum_{k=1}^{n} \xi_k$ 的特征函数

$$f_{\sum_{k=1}^{n} \xi_k}(t) = \prod_{k=1}^{n} f_{\xi_k}(t) = \prod_{k=1}^{n} \left(e^{ita_k - \frac{\sigma_k^2 t^2}{2}}\right) = e^{it\left(\sum_{k=1}^{n} a_k\right) + \frac{\left(\sum_{k=1}^{n} \sigma_k^2\right)t^2}{2}}. \tag{22}$$

但(22)是分布 $N\left(\sum_{k=1}^{n} a_k, \sum_{k=1}^{n} \sigma_k^2\right)$ 的特征函数，根据唯一性定理，只有分布 $N\left(\sum_{k=1}^{n} a_k, \sum_{k=1}^{n} \sigma_k^2\right)$ 才有特征函数(22)，所以 $\sum_{k=1}^{n} \xi_k \sim N\left(\sum_{k=1}^{n} a_k, \sum_{k=1}^{n} \sigma_k^2\right)$. ▌

【例 9】 ξ_k，$k=1, 2, \cdots, r$ 独立，分别是自由度为 $n_1, n_2, \cdots, n_r$ 的 χ^2 变量，试求 $\sum_{k=1}^{r} \xi_k$ 的分布.

解: 由性质 5 及例 6，$\sum_{k=1}^{r} \xi_k$ 的特征函数

$$f_{\sum_{k=1}^{r} \xi_k}(t) = \prod_{k=1}^{r} f_{\xi_k}(t) = \prod_{k=1}^{r} (1-2it)^{-\frac{n_k}{2}} = (1-2it)^{-\frac{\sum_{k=1}^{r} n_k}{2}}. \tag{23}$$

由例 6 及唯一性定理，$\sum_{k=1}^{r} \xi_k$ 服从自由度为 $\sum_{k=1}^{r} n_k$ 的 χ^2 分布.

从例 8、例 9 可以看到，在有了特征函数这个工具后，求随机变量的分布的问题有时变得很简单，这一点在下一章中还会见到.

在结束这一节时我们向读者指出，对于多维随机变量，特征函数也有相应的内容. 多维随机变量 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的特征函数指

$$Ee^{i(t_1\xi_1+t_2\xi_2+\cdots+t_n\xi_n)}$$

是实向量$(t_1, t_2, \cdots, t_n)$的函数，它也有类似于一维随机变量特征函数的性质及唯一定理. 有兴趣的读者可以参看有关书籍.

习 题 五

1. ξ 在$[-a, a]$上均匀分布，试求 ξ 的特征函数.
2. 试求二项分布 $b(n, p)$的特征函数.
3. ξ 服从参数为 1 的指数分布，即有概率分布密度

$$p(x)=\begin{cases} e^{-x}, & x\geqslant 0;\\ 0, & x<0. \end{cases}$$

试求 ξ 的特征函数.

4. η 服从指数分布

$$p(x)=\begin{cases} be^{-bx}, & x\geqslant 0;\\ 0, & x<0. \end{cases}$$

试用上题结果求 η 的特征函数.

5. 利用特征函数求上题中的 $E\eta$, $D\eta$.
6. 若特征函数是 $\frac{e^{it}(1-e^{nit})}{n(1-e^{it})}$，试证对应的随机变量有分布列

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

7. ξ 有特征函数

$$f(t)=\frac{1}{16}(1+2e^{it}+e^{2it})^2,$$

试求 (1) $E\xi$, (2) $D\xi$, (3) $P(1<\xi<4)$, (4) $P(\xi>2)$.

8. 证明:有限个独立的，服从泊松分布的随机变量之和仍服从泊松分布.
9. ξ, η 分别是自由度为 m 和 n 的 χ^2 变量. 已知 $n>m$, $\eta-\xi$ 与 ξ 独立，试证明 $\eta-\xi$ 是自由度为 $n-m$ 的 χ^2 变量.

补 充 题

1. ξ 是取非负整数值的离散型随机变量，有分布列

$$P(\xi=k)=p_k,\quad k=0,1,2,\cdots,$$

以 $p_k(k=0,1,2,\cdots)$ 为系数的幂级数 $\sum_{k=0}^{+\infty}P_kt^k$ 收敛半径 $r\geqslant1$；称函数 $P_\xi(t)=\sum_{k=0}^{+\infty}p_kt^k$ 为 ξ 的(概率)母函数，当 $r>1$ 时，有

$$\frac{d}{dt}P_\xi(t)\Big|_{t=1}=\sum_{k=1}^{+\infty}kp_k=E\xi,$$

$$\frac{d^2}{dt^2}P_\xi(t)\Big|_{t=1}=\sum_{k=2}^{+\infty}k(k-1)p_k=E\xi(\xi-1)=E\xi^2-E\xi$$

等等．试求下列分布的母函数及数学期望、方差：

（i）二项分布：$0<p<1$，$q=1-p$，

$$P(\xi=k)=C_n^kp^kq^{n-k},\quad k=0,1,2,\cdots,n.$$

（ii）泊松分布：$\lambda>0$，

$$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots.$$

（iii）负二项分布*)：$0<p<1$，$q=1-p$，

$$P(\xi=k)=C_{r+k-1}^kp^rq^k,\quad k=0,1,2,\cdots.$$

2. 在一台仪器上进行重复试验，直到仪器发生故障为止．设在每次试验中发生故障的概率都等于 $\frac{1}{10}$，无故障试验的次数为 ξ，试求 $E\xi$．

3. μ 是事件 A 在 n 次独立试验中出现的次数，$P(A)=p$，试求
(i) $E\mu^3$，(ii) $E\mu^4$．

4. $\xi_1,\xi_2,\cdots,\xi_n$ 独立，都在 $[a,b]$ 上均匀分布，$\eta_1=\min\{\xi_1,\cdots,\xi_n\}$，$\eta_n=\max\{\xi_1,\cdots,\xi_n\}$，试求
(i) $E\eta_1$，$E\eta_n$；(ii) $D\eta_1$，$D\eta_n$；(iii) $D(\eta_1+\eta_n)$．

5. 袋中有 k 号的球 k 只，$k=1,2,\cdots,n$．从中摸出一球，求所得号码

*) 在一个独立试验序列中，每次成功的概率为 p，$\xi+r$ 表示第 r 次成功的试验序号．

的数学期望和方差.

6. 求超几何分布的数学期望和方差.
7. ξ, η 都是只能取两个值的随机变量,试证: 若 ξ, η 无关, 则 ξ, η 独立.
8. $\xi_i \sim N(a, \sigma^2)$, $i=1, 2$ 独立,证明 $\xi_1+\xi_2$, $\xi_1-\xi_2$ 独立.

第三章小结

随机变量或其分布律的数字特征能刻划分布律的某些特点. 对某些类型的分布律来说, 知道了它的某些数字特征也就知道了分布律. 在第二章小结中已经说过: 研究分布律及其数字特征是本书的中心内容.

数学期望与方差是读者应当注意的两个数字特征.由1.1 (3)、(6)定义的数值叫随机变量的数学期望,它在一定意义下表示随机变量的平均值. 方差是刻划概率分布集中(或分散)程度的一个数字特征. 要熟悉数学期望和方差的性质, 记住常见分布(如正态分布,泊松分布)的数学期望和方差.

车贝谢夫不等式是一个重要的不等式, 在第四章将看到它的应用.

特征函数与分布律之间的一一对应关系使特征函数成为研究分布律的一个工具, 特别是研究独立随机变量和的分布的重要工具. 本章在微积分的基础上讨论特征函数. 如使用进一步的数学工具可以使讨论简化.

在本书中与多维正态分布有关的内容是分散在第二章, 第三章中讲的, 有兴趣的读者可以将这些内容系统整理一下.

第四章

大数定律及中心极限定理

$\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是随机变量序列. 在这一章主要讲两类极限定理: 一类是关于随机变量序列

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k, \quad n=1, 2, \cdots$$

"收敛"于零的极限定理, 叫大数定律; 一类是关于随机变量序列

$$\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}}, \quad n=1, 2, \cdots$$

对应的分布函数序列收敛于标准正态分布函数的定理, 叫中心极限定理.

随着这部分内容的展开, 读者将会看到研究这两类极限定理的重要意义.

第一节　依概率收敛与几乎处处收敛

随机变量序列收敛的概念是一个新的概念, 它有各种意义下的收敛, 这里只介绍几乎处处收敛与依概率收敛两个概念.

1.1 几乎处处收敛

设 $\xi(\omega), \xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega), \cdots$ 是定义在概率空间

$(\Omega, \mathscr{F}, P)$上的随机变量，对于任意指定的 $\omega_0\in\Omega$，

$$\xi_1(\omega_0),\ \xi_2(\omega_0),\ \cdots,\ \xi_n(\omega_0),\ \cdots \tag{1}$$

是一个普通的数列，$\xi(\omega_0)$是一个普通的数．数列(1)或收敛于$\xi(\omega_0)$，或不收敛于$\xi(\omega_0)$．考虑Ω中使 $\lim\limits_{n\to+\infty}\xi_n(\omega)=\xi(\omega)$ 成立的子集

$$\Omega'=\{\omega \mid \omega\in\Omega,\ \lim_{n\to+\infty}\xi_n(\omega)=\xi(\omega)\},$$

或简单地记 $\Omega'=\{\lim\limits_{n\to+\infty}\xi_n=\xi\}$．

定义 $\xi,\ \xi_1,\ \xi_2,\ \cdots,\ \xi_n,\ \cdots$ 都是随机变量，若

$$P(\Omega')=P(\lim_{n\to+\infty}\xi_n=\xi)=1,$$

就称随机变量序列 $\{\xi_n\}$ 几乎处处收敛于随机变量ξ，记作 $\xi_n\xrightarrow{a.e(P)}\xi$．

在谈Ω'的概率时，自然应当说明$\Omega'\in\mathscr{F}$．我们将证明

$$\Omega'=\bigcap_{m=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{|\xi_n-\xi|<\frac{1}{m}\right\}. \tag{2}$$

由于

$$\left\{|\xi_n-\xi|<\frac{1}{m}\right\}=\bigcup_{r=1}^{\infty}\left\{-\frac{1}{m}+\frac{1}{r}\leqslant\xi_n-\xi<\frac{1}{m}\right\}\in\mathscr{F},$$

在证明了(2)后立刻可以得到$\Omega'\in\mathscr{F}$．

等式(2)的证明：

$$\omega\in\{\lim_{n\to+\infty}\xi_n=\xi\}\Leftrightarrow \lim_{n\to+\infty}\xi_n(\omega)=\xi(\omega)$$

$\Leftrightarrow$ 对任意自然数 m，存在自然数 N，当 $n\geqslant N$ 时，

$$|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}$$

$\Leftrightarrow$ 对任意自然数 m，存在自然数 N，当 $n\geqslant N$ 时，

$$\omega\in\left\{|\xi_n-\xi|<\frac{1}{m}\right\}$$

$\Leftrightarrow$ 对任意自然数 m, 存在自然数 N,

$$\omega\in\bigcap_{n=N}^{\infty}\left\{|\xi_n-\xi|<\frac{1}{m}\right\}$$

$\Leftrightarrow$ 对任意自然数 m,

$$\omega\in\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{|\xi_n-\xi|<\frac{1}{m}\right\}$$

$$\Leftrightarrow\qquad \omega\in\bigcap_{m=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{|\xi_n-\xi|<\frac{1}{m}\right\}. \qquad ▌$$

由定义及(2)可以看到 $\xi_n\xrightarrow{a.e(P)}\xi$ 和 $(\xi_n-\xi)\xrightarrow{a.e(P)}0$ 是等价的.

1.2 依概率收敛

依概率收敛是和几乎处处收敛有区别的另一个收敛概念. $\xi,\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是随机变量. 对任意给定的实数 $\varepsilon>0$, 随机事件序列

$$A_{n,\varepsilon}=\{|\xi_n-\xi|<\varepsilon\},\quad n=1,2,\cdots$$

对应的概率 $P(A_{n,\varepsilon})$, $n=1,2,\cdots$,是一个普通的数列.

定义 $\xi,\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 都是随机变量, 若对任意实数 $\varepsilon>0$,

$$\lim_{n\to+\infty}P\{|\xi_n-\xi|<\varepsilon\}=1,$$

就称随机变量序列 $\{\xi_n\}$ 依概率收敛于 ξ, 记作 $\xi_n\xrightarrow{P}\xi$.

显然, $\xi_n\xrightarrow{P}\xi$ 与 $(\xi_n-\xi)\xrightarrow{P}0$ 是等价的.

$\xi_n\xrightarrow{P}\xi$ 意味着只要 n 充分大, ξ_n 和 ξ 取的值可以以很大的概率非常接近.

当 $\xi_n\xrightarrow{a.e(P)}\xi$ 时,必有 $\xi_n\xrightarrow{P}\xi$.

这是因为对任何 $\varepsilon>0$, 有自然数 m, $\frac{1}{m}<\varepsilon$,

$$\begin{aligned}\lim_{n\to+\infty}P(|\xi_n-\xi|<\varepsilon)&\geqslant\lim_{n\to+\infty}P\left(|\xi_n-\xi|<\frac{1}{m}\right)\\&\geqslant\lim_{n\to+\infty}P\left(\bigcap_{k=n}^{\infty}\left[|\xi_k-\xi|<\frac{1}{m}\right]\right)^{*)}\\&=P\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left[|\xi_n-\xi|<\frac{1}{m}\right]\right)\\&\geqslant P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left[|\xi_n-\xi|<\frac{1}{m}\right]\right)=1.\end{aligned}$$

但是 $\xi_n\xrightarrow{P}\xi$ 时, 不一定有 $\xi_n\xrightarrow{a.s.(P)}\xi$. 例如, 取 $\Omega=(0,1]$, $\mathscr{F}$ 是包含 $(0,1]$ 中一切左开右闭区间的事件域, P 是定义在 $\mathscr{F}$ 上的概率, 满足 $P((a,b])=b-a$. 定义随机变量

$$\eta_{11}(\omega)\equiv1;$$

$$\eta_{21}(\omega)=\begin{cases}1, & \text{当 }\omega\in\left(0,\frac{1}{2}\right],\\0, & \text{当 }\omega\in\left(\frac{1}{2},1\right];\end{cases}$$

$$\eta_{22}(\omega)=\begin{cases}0, & \text{当 }\omega\in\left(0,\frac{1}{2}\right],\\1, & \text{当 }\omega\in\left(\frac{1}{2},1\right];\end{cases}$$

…………………………

一般地, 将 $(0,1]$ 分成 k 个等长的区间, 令

$$\eta_{ki}(\omega)=\begin{cases}1, & \text{当 }\omega\in\left(\frac{i-1}{k},\frac{i}{k}\right],\\0, & \text{当 }\omega\notin\left(\frac{i-1}{k},\frac{i}{k}\right];\end{cases}$$

---

*) 令 $A_n=\bigcap_{k=n}^{\infty}\left[|\xi_k-\xi|<\frac{1}{m}\right]$, $n=1,2,\cdots$, 则 $A_1\subseteq A_2\subseteq\cdots\subseteq A_n\subseteq\cdots$ 利用第一章第五节性质 9 即可得证.

$$(i=1,\ 2,\ \cdots,\ k;\quad k=1,\ 2,\ \cdots)$$

则对任意给定的 $\varepsilon>0$, 有

$$P(|\eta_{ki}|\geqslant\varepsilon)\leqslant\frac{1}{k}. \tag{3}$$

考虑随机变量序列

$$\xi_1(\omega)=\eta_{11}(\omega),\quad \xi_2(\omega)=\eta_{21}(\omega),$$
$$\xi_3(\omega)=\eta_{22}(\omega),\quad \xi_4(\omega)=\eta_{31}(\omega),$$
$$\xi_5(\omega)=\eta_{32}(\omega),\ \cdots.$$

由(3)知

$$\lim_{n\to+\infty}P(|\xi_n|\geqslant\varepsilon)=0,$$

即 $\xi_n\xrightarrow{P}0$. 但对任一指定的 $\omega_0\in\Omega$, 数列 $\{\xi_n(\omega_0)\}$中有无穷多个1, 无穷多个0, 由柯西收敛原理, $\{\xi_n(\omega_0)\}$发散, 所以 $\{\xi_n\}$ 不几乎处处收敛于零.

依概率收敛有下面的重要性质:

定理 1 $\{\xi_n\}$, $\{\eta_n\}$是两个随机变量序列, a、b是常数. 若 $\xi_n\xrightarrow{P}a$, $\eta_n\xrightarrow{P}b$, 函数 $g(x,\ y)$在点$(a,\ b)$连续,则

$$g(\xi_n,\ \eta_n)\xrightarrow{P}g(a,\ b).$$

【证】 因为 $g(x,y)$在点(a,b)连续,对任意给定的 $\varepsilon>0$, 存在 $\delta>0$, 当$(x-a)^2+(y-b)^2<\delta$ 时,

$$|g(x,\ y)-g(a,\ b)|<\varepsilon.$$

用 $A_{n\delta}$ 表示事件$\left(|\xi_n-a|<\sqrt{\frac{\delta}{2}}\right)$, $B_{n\delta}$ 表示事件

$$\left(|\eta_n-b|<\sqrt{\frac{\delta}{2}}\right),\quad n=1,\ 2,\ \cdots,$$

则有 $\lim\limits_{n\to+\infty}P(A_{n\delta})=1,\quad \lim\limits_{n\to+\infty}P(B_{n\delta})=1.$

因为 $P(A_{n\delta}\cup B_{n\delta})\geqslant P(A_{n\delta}),$

而有 $\lim\limits_{n\to+\infty}P(A_{n\delta}\cup B_{n\delta})=1.$

又 $$\begin{aligned}&P(|g(\xi_n,\eta_n)-g(a,b)|<\varepsilon)\\&\geqslant P((\xi_n-a)^2+(\eta_n-b)^2<\delta)\\&\geqslant P\left(|\xi_n-a|<\sqrt{\frac{\delta}{2}},\ |\eta_n-b|<\sqrt{\frac{\delta}{2}}\right)\\&=P(A_{n\delta}\cap B_{n\delta})\\&=P(A_{n\delta})+P(B_{n\delta})-P(A_{n\delta}\cup B_{n\delta}),\end{aligned}$$

所以 $$\lim_{n\to+\infty}P(|g(\xi_n,\eta_n)-g(a,b)|<\varepsilon)=1. \quad ▌$$

分别取 $g(x,y)$ 为 $x+y$, $x-y$, xy, $\frac{x}{y}$ 则有

推论 1 若 $\xi_n\xrightarrow{P}a$, $\eta_n\xrightarrow{P}b$, 则 $\xi_n\pm\eta_n\xrightarrow{P}a\pm b$, $\xi_n\eta_n\xrightarrow{P}ab$; 若 $b\neq0$, 还有 $\frac{\xi_n}{\eta_n}\xrightarrow{P}\frac{a}{b}$.

这个定理可以推广到任意有限个随机变量序列的情况, 读者可以试证.

第二节 大数定律与强大数定律

$\xi_1, \xi_2, \cdots, \xi_n, \cdots$是随机变量序列.

若 $\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\xrightarrow{P}0$, 则称$\{\xi_k\}$服从大数定律.

若$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\xrightarrow{a.e(P)}0$, 则称 $\{\xi_k\}$ 服从强大数定律.

定理 2 (马尔科夫大数定律) $\{\xi_k\}$是随机变量序列, 若

$$\lim_{n\to+\infty}\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=0, \tag{4}$$

则 $\{\xi_k\}$ 服从大数定律.

【证】 由(4), 当 n 充分大时, $D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)<+\infty$. 对

任意给定的 $\varepsilon>0$，由车贝谢夫不等式有

$$0\leqslant P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|\geqslant\varepsilon\right)$$

$$\leqslant\frac{D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)}{\varepsilon^2}=\frac{D\left(\sum_{k=1}^{n}\xi_k\right)}{n^2}\cdot\frac{1}{\varepsilon^2}.$$

所以 $$\lim_{n\to+\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|\geqslant\varepsilon\right)=0,$$

$$\lim_{n\to+\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|<\varepsilon\right)=1.\quad ▮$$

推论2(车贝谢夫大数定律) $\{\xi_k\}$是独立随机变量序列，若有常数 C 使 $D\xi_k\leqslant C$，$k=1, 2, \cdots$，则 $\{\xi_k\}$服从大数定律.

【证】 $$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D\xi_k\leqslant\frac{nC}{n^2},$$

所以 $$\lim_{n\to+\infty}\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=0,$$

由定理1得证. ▮

推论3(贝努里大数定律) $\{\xi_k\}$是独立随机变量序列，ξ_k 有分布列

$$\begin{array}{c|cc}\xi_k & 0 & 1\\ \hline P_{\xi_k} & 1-p & p\end{array}\quad k=1, 2, \cdots, \tag{5}$$

则 $\{\xi_n\}$ 服从大数定律，具体地说有

$$\frac{\xi_1+\cdots+\xi_n}{n}\xrightarrow{P}p, \tag{6}$$

即对任意给定的正数 ε，

$$\lim_{n\to+\infty}P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|<\varepsilon\right)=1.$$

【证】 $D\xi_k=p(1-p)\leqslant1,\quad k=1, 2, \cdots,$

由推论2

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\xrightarrow{P}0.$$

$E\xi_k=p$, $k=1, 2, \cdots$，代入上式，即得(6). ∎

在第一章引言中，描述重复试验中事件 A 出现的规律时说过，事件 A 的频率具有稳定性，它经常与一固定的数值(即 A 的概率 p)接近，而且随着试验重复次数的增加，这种现象愈加显著. 现在，我们可以利用大数定律对上述事实给出更加确切的解释. 设事件 A 的概率为 p, ξ_k 是第 k 次独立重复试验中事件 A 出现的次数，则 ξ_k 有分布列(5)，从而利用贝努里大数定律知

$$\frac{\xi_1+\cdots+\xi_n}{n}\xrightarrow{P}p.$$

这就是说，重复试验中事件 A 出现的频率依概率收敛于它的概率 p.

定理2证明了在方差满足一定条件的情况下随机变量序列 $\{\xi_k\}$ 是服从大数定律的. 另一个常见的重要情况是 $\{\xi_k\}$ 独立，并且有相同的分布. 这时如果 ξ_k 的方差存在，那么 $\{\xi_k\}$ 就具备了推论2的条件，毫无疑问，$\{\xi_k\}$ 也服从大数定律. 需要进一步考虑的是 ξ_k 的方差不存在时，$\{\xi_k\}$ 是否还服从大数定律. 下面将要证明的辛钦定理回答了这个问题.

定理3 (辛钦(Хинчин)大数定律) $\{\xi_k\}$ 是独立同分布的随机变量序列，若数学期望存在($E\xi_k=a$)，则 $\{\xi_k\}$ 服从大数定律.

以下只对 ξ_k 是连续型随机变量的情况给以证明. 当 ξ_k 是离散型随机变量时，证明是类似的.

【证】 设 ξ_k 有概率分布密度 $p(x)$，根据依概率收敛的定义我们应证明：对任意给定的 $\varepsilon>0$, $\varepsilon'>0$, 存在 $N>0$,

当 $n>N$ 时

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a\right|\geqslant\varepsilon\right)<\varepsilon'.$$

由于 $E\xi_k$ 存在，积分 $\int_{-\infty}^{+\infty}|x|p(x)dx$ 收敛，设积分值为 b，再设 $\delta=\min\left\{\frac{\varepsilon'\varepsilon^2}{8b},\ \varepsilon'\right\}$. 因为当 $n\to+\infty$ 时，

$$a_n=\int_{-n\delta}^{n\delta}xp(x)dx\to\int_{-\infty}^{+\infty}xp(x)dx=a, \tag{7}$$

$$\int_{|x|\geqslant n\delta}|x|p(x)dx\to0, \tag{8}$$

所以有 $N>0$，当 $n>N$ 时

$$|a_n-a|<\frac{\varepsilon}{2},$$

$$\int_{|x|\geqslant n\delta}|x|p(x)dx<\frac{\delta^2}{2}. \tag{9}$$

对大于 N 的任何一个 n，令

$$\eta_k=f(\xi_k)=\begin{cases}\xi_k, & \text{当}\ |\xi_k|<n\delta,\\ 0, & \text{当}\ |\xi_k|\geqslant n\delta,\end{cases}\quad k=1,\ 2,\ \cdots,n, \tag{10}$$

这里

$$f(x)=\begin{cases}x, & \text{当}\ |x|<n\delta;\\ 0, & \text{当}\ |x|\geqslant n\delta.\end{cases}$$

则 $\eta_1,\ \eta_2,\ \cdots,\ \eta_n$ 独立，并且对任何 $1\leqslant k\leqslant n$ 有

$$E\eta_k=Ef(\xi_k)=\int_{-\infty}^{+\infty}f(x)p(x)dx=\int_{-n\delta}^{n\delta}xp(x)dx=a_n,$$

$$D\eta_k=E\eta_k^2-(E\eta_k)^2=\int_{-n\delta}^{n\delta}x^2p(x)dx-(E\eta_k)^2$$

$$\leqslant n\delta\int_{-n\delta}^{n\delta}|x|p(x)dx\leqslant nb\delta.$$

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a\right|\geqslant\varepsilon\right)$$
$$\leqslant P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a_n\right|+|a_n-a|\geqslant\varepsilon\right)$$
$$\leqslant P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a_n\right|\geqslant\frac{\varepsilon}{2}\right)$$
$$=P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a_n\right|\geqslant\frac{\varepsilon}{2},\ \bigcap_{k=1}^{n}\{\xi_k=\eta_k\}\right)$$
$$+P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a_n\right|\geqslant\frac{\varepsilon}{2},\ \bigcup_{k=1}^{n}\{\xi_k\neq\eta_k\}\right)$$
$$\leqslant P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\eta_k-a_n\right|\geqslant\frac{\varepsilon}{2}\right)+\sum_{k=1}^{n}P(\xi_k\neq\eta_k), \quad (11)$$

由车贝谢夫不等式

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\eta_k-a_n\right|\geqslant\frac{\varepsilon}{2}\right)\leqslant\frac{4\sum\limits_{k=1}^{n}D\eta_k}{n^2\varepsilon^2}$$
$$\leqslant\frac{4n^2b\delta}{n^2\varepsilon^2}\leqslant\frac{\varepsilon'}{2}, \quad (12)$$

又由(9),(10)

$$\sum_{k=1}^{n}P(\xi_k\neq\eta_k)=\sum_{k=1}^{n}P(|\xi_k|\geqslant n\delta)$$
$$=\sum_{k=1}^{n}\int_{|x|\geqslant n\delta}p(x)dx\leqslant\sum_{k=1}^{n}\frac{1}{n\delta}\int_{|x|\geqslant n\delta}|x|p(x)dx$$
$$<\frac{1}{\delta}\cdot\frac{\delta^2}{2}=\frac{\delta}{2}\leqslant\frac{\varepsilon'}{2}. \quad (13)$$

将(12)、(13)代回(11),得证. ▌

在第三章定义随机变量的数学期望时,我们对数学期望的含义作了一些说明,现在可以利用大数定律作出更加确切的解释. 设随机变量 ξ 有数学期望 a. 用 ξ_k 表示在重复试验中 ξ 取的第 k 个值,则可以认为 ξ_k 是随机变量,且 $\{\xi_k\}$ 是独立随机变量序列,每一个 ξ_k 和 ξ 有相同的分布. 由定理3,

对任意给定的 $\varepsilon>0$ 就有

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a\right|<\varepsilon\right)=1. \qquad (14)$$

(14)式说明在第三章定义中定义的 $E\xi=a$ 是这样一个数，只要 n 充分大，ξ 取的 n 个值的平均值可以以很大的概率和这个数充分接近. 在第三章我们曾经用只取有限个值的离散型随机变量解释过这层意思，不过那时我们用了"稳定"这样一个不确切的概念，现在用(14)说明 ξ 重复取的 n 个值的平均值 $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ 和 $E\xi=a$ 的关系就更加确切了.

下面用几个简单的例子说明大数定律的应用.

1．人们在进行精密测量时，为了减少随机误差，往往重复测量多次然后取算术平均值. 这样做能达到预想的目的吗？假设你用一架精密天平测量一个实重为 a 的东西. 由于测量环境中随机因素的干扰，多次测量的结果为 $a+\delta_1$, $a+\delta_2$, …，这里 δ_1, δ_2, …表示随机误差，n 次测量结果的平均值就是 $a+\frac{1}{n}\sum_{k=1}^{n}\delta_k$. 视 δ_1, δ_2, …为随机变量，可以认为它们是独立的，同分布的，并且 $E\delta_1=0$(否则就表示含有系统误差). 由定理 3

$$\frac{1}{n}\sum_{k=1}^{n}\delta_k \xrightarrow{P} 0.$$

这说明当测量次数比较多时，平均值 $a+\frac{1}{n}\sum_{k=1}^{n}\delta_k$ 和真值 a 的差 $\frac{1}{n}\sum_{k=1}^{n}\delta_k$ 能以比较大的概率取接近于 0 的值.

2．根据定义，要想知道一个随机变量 ξ 的数学期望 $E\xi$，方差 $D\xi$，必须知道 ξ 的分布律. 然而在实际问题中往往是在不知道 ξ 分布律的情况下希望知道 $E\xi$ 和 $D\xi$. 例如，在某

种工艺条件下生产一种零件，由于随机因素的干扰，某一项指标（如长度）是随机变量ξ. 为了检查零件的质量，不一定要知道ξ的分布，只要能对$E\xi$, $D\xi$做出估计就行了. 我们可以随机抽取一些零件，测量其长度得$x_1, x_2, \cdots, x_n$, 用平均值$\frac{1}{n}\sum_{k=1}^{n} x_k$作为$E\xi$的估计值. 因为抽取的第$k$个零件的长度是随机变量，用$\xi_k$表示，由定理3，当$n$比较大时，$\frac{1}{n}\sum_{k=1}^{n}\xi_k$能以比较大的概率取接近于$E\xi$的值，而$\frac{1}{n}\sum_{k=1}^{n} x_k$就是$\frac{1}{n}\sum_{k=1}^{n}\xi_k$取的一个值，所以，用$\frac{1}{n}\sum_{k=1}^{n} x_k$作为$E\xi$的估计值，当$n$比较大时，一般会比较接近于$E\xi$.

为了估计$D\xi$, 由于$D\xi=E\xi^2-(E\xi)^2$, 可以先分别估计$E\xi^2$和$(E\xi)^2$. $\{\xi_k\}$是独立同分布随机变量序列，因而$\{\xi_k^2\}$也是独立同分布随机变量序列，每一个ξ_k^2和ξ^2有相同的分布，故

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2 \xrightarrow{P} E\xi^2,$$

可以用$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2$取的值做为$E\xi^2$的估计值. 又

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{P} E\xi,$$

由定理1

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2-\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)^2 \xrightarrow{P} E\xi^2-(E\xi)^2=D\xi,$$

我们可以用$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2-\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)^2$取的值作为$D\xi$的估计值，具体说，就是用

$$\frac{1}{n}\sum_{k=1}^{n}x_k^2-\left(\frac{1}{n}\sum_{k=1}^{n}x_k\right)^2$$

作为 $D\xi$ 的估计值.

3. 计算定积分 $J=\int_a^b g(x)dx$ 的近似值.

令

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & \text{当 } a\leqslant x\leqslant b,\\ 0, & \text{当 } x<a \text{ 或 } x>b,\end{cases} \tag{15}$$

则

$$J=(b-a)\int_{-\infty}^{+\infty} g(x)p(x)dx.$$

(15)是在 $[a, b]$ 上均匀分布的随机变量的概率分布密度, 由第三章 1.1 定理, $g(\xi)$ 的数学期望就是 $\dfrac{J}{(b-a)}$, 由定理 3, 只要能得到独立的、每一个都在 $[a, b]$ 上均匀分布的随机变量序列 $\{\xi_k\}$ 取的值 $\{x_k\}$, 由于

$$\frac{1}{n}\sum_{k=1}^{n} g(\xi_k)\xrightarrow{P}\frac{J}{b-a},$$

就可以用 $(b-a)\sum\limits_{k=1}^{n}\dfrac{g(x_k)}{n}$ 作为 J 的近似值. 数列 $\{x_k\}$ 是可以由计算机产生的.

关于强大数定律我们只证明定理 4, 更一般的情形只列出定理 5 而不证明, 有兴趣的读者可参考严士健等《测度论与概率论基础》.

定理 4(波雷尔强大数定律) $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 独立, 有相同的分布列 $\begin{pmatrix} 0 & 1 \\ 1-p & p\end{pmatrix}$, 则 $\{\xi_k\}$ 服从强大数定律.

【证】 需要证明的是

$$P\left(\bigcup_{m=1}^{\infty}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right\}\right)=0,$$

或对任何正整数 m

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right\}\right)=0.$$

但

$$\begin{aligned}
&P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right\}\right)\\
&=\lim_{N\to+\infty}P\left(\bigcup_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right\}\right)\\
&\leqslant\lim_{N\to+\infty}\sum_{n=N}^{\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right),
\end{aligned}$$

所以只要能证明级数 $\sum_{n=1}^{\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right)$ 收敛就行了.

由第三章1.2性质8,

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right)\leqslant E\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|^4\Big/\left(\frac{1}{m}\right)^4\\
&=\frac{m^4}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E(\xi_i-p)(\xi_j-p)(\xi_k-p)(\xi_l-p)\\
&=\frac{m^4}{n^4}\left[\sum_{i=1}^{n}E(\xi_i-p)^4+\sum_{1\leqslant i<j\leqslant n}C_4^2E(\xi_i-E\xi_i)^2(\xi_j-E\xi_j)^2\right]\\
&=\frac{m^4}{n^4}\{n[p^4(1-p)+(1-p)^4p]+3n(n-1)p^2(1-p)^2\}\\
&\leqslant\frac{m^4}{n^4}[2n+3n(n-1)]\\
&\leqslant\frac{m^4}{n^4}\cdot3n^2=\frac{3m^4}{n^2},
\end{aligned}$$

所以

$$\sum_{n=1}^{\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-p\right|\geqslant\frac{1}{m}\right)<+\infty. \quad \blacksquare$$

定理5(柯尔莫哥洛夫强大数定律) 在定理3的条件下

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k\xrightarrow{a.e.(P)}a.$$

第三节　中心极限定理

本节定理的证明过程中需要用特征函数的两个极限定理，这里只列出定理而不证明*).

正极限定理　如果分布函数序列 $F_1(x)$, $F_2(x)$, …, $F_n(x)$…在分布函数 $F(x)$的连续点上收敛于分布函数 $F(x)$，则对应的特征函数序列 $f_1(t)$, $f_2(t)$, …, $f_n(t)$, …收敛于 $F(x)$的特征函数 $f(t)$，并且在 t 的每一个有限区间内一致收敛于 $f(x)$.

逆极限定理(连续性定理)　如果对应于分布函数序列

$$F_1(x),\ F_2(x),\ \cdots,\ F_n(x),\ \cdots \tag{16}$$

的特征函数序列

$$f_1(t),\ f_2(t),\ \cdots,\ f_n(t),\ \cdots \tag{17}$$

收敛于连续函数 $f(t)$，则有一个分布函数 $F(x)$，在其连续点上，(16)收敛于 $F(x)$. (由正极限定理，$F(x)$的特征函数就是 $f(t)$.)

作为一个例题，这里用特征函数的极限定理重新证明定理3.

$\{\xi_k\}$ 是独立同分布的随机变量序列，数学期望 $E\xi_k=a$ 存在. 因为诸 ξ_k 分布相同，应有相同的特征函数，设为 $f(t)$. 由特征函数的性质2、3，$\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k=\sum\limits_{k=1}^{n}\frac{\xi_k}{n}$ 的特征函数是 $\left[f\left(\frac{t}{n}\right)\right]^n$，由性质6，在 $t=0$ 的附近

$$f(t)=1+\frac{f'(0)}{1!}t+o(t)=1+iat+o(t).$$

*) 证明可参看严士健等《测度论与概率论基础》.

因此，对任一固定的 t 有

$$\lim_{n\to+\infty}\left[f\left(\frac{t}{n}\right)\right]^n=\lim_{n\to+\infty}\left[1+\frac{iat}{n}+o\left(\frac{t}{n}\right)\right]^n=e^{iat},$$

e^{iat} 是连续函数，并且是分布函数

$$F(x)=\begin{cases}1, & 当\ x<a;\\ 0, & 当\ x\leqslant a\end{cases}$$

的特征函数. 由特征函数的唯一性定理及逆极限定理, $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ 的分布函数当 $n\to+\infty$ 时在 $F(x)$ 的连续点上以 $F(x)$ 为极限，故对任何 $\varepsilon>0$ 有

$$\begin{aligned}&\lim_{n\to+\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a\right|<\varepsilon\right)\\&=\lim_{n\to+\infty}P\left(a-\varepsilon<\frac{1}{n}\sum_{k=1}^{n}\xi_k<a+\varepsilon\right)\\&\geqslant\lim_{n\to+\infty}P\left(a-\frac{\varepsilon}{2}\leqslant\frac{1}{n}\sum_{k=1}^{n}\xi_k<a+\varepsilon\right)\\&=\lim_{n\to+\infty}\left[F_n(a+\varepsilon)-F_n\left(a-\frac{\varepsilon}{2}\right)\right]=1.\quad\blacksquare\end{aligned}$$

定理 6 $\{\xi_k\}$ 是独立同分布的随机变量序列, $E\xi_1=a$, $D\xi_1=\sigma^2$(有限)，则对任何实数 x,

$$\lim_{n\to+\infty}P\left(\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}}<x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt. \quad (18)$$

【证】 设 ξ_k-a 的特征函数为 $f(t)$，则由特征函数性质 2, 5,

$$\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}}=\sum_{k=1}^{n}\frac{\xi_k-a}{\sqrt{n}\,\sigma}$$

的特征函数为$\left[f\left(\frac{t}{\sqrt{n}\,\sigma}\right)\right]^n$. 由于$\xi_k-a$的二阶矩存在,由特征函数性质6, $f(t)$一定可微分两次. 将$f(t)$在$t=0$附近展成泰勒级数

$$f(t)=f(0)+\frac{f'(0)}{1!}t+\frac{f''(0)}{2!}t^2+o(t^2),$$

因为 $$f(0)=1,\quad f'(0)=iE(\xi_k-a)=0,$$

$$f''(0)=i^2E(\xi_k-a)^2=-\sigma^2,$$

故 $$f(t)=1-\frac{\sigma^2t^2}{2}+o(t^2).$$

对任何指定的t

$$\lim_{n\to+\infty}\left[f\left(\frac{t}{\sqrt{n}\,\sigma}\right)\right]^n=\lim_{n\to+\infty}\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n=e^{-\frac{t^2}{2}},$$

$e^{-\frac{t^2}{2}}$是$N(0,1)$的特征函数,由逆极限定理,(18)式得证. ∎

推论4 $\{\xi_k\}$是独立随机变量序列, ξ_k, $k=1, 2, \cdots$有相同的分布列$\begin{pmatrix}0 & 1\\ 1-p & p\end{pmatrix}$, 则对任何实数$x$

$$\lim_{n\to+\infty}P\left(\frac{\sum\limits_{k=1}^{n}\xi_k-np}{\sqrt{np(1-p)}}<x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt. \tag{19}$$

【例1】 某工厂生产的产品废品率是0.005, 试求任意抽取的10000件产品中废品数不多于70件的概率p.

解: 这是一个n次独立试验中事件的概率计算问题. 用第一章第八节的方法

$$p=\sum_{k=0}^{70}C_{10000}^{k}(0.005)^k(0.995)^{10000-k},$$

但直接计算计算量太大了.

用ξ_k表示取来的第k件产品中的废品数, $k=1, 2, \cdots$, 则$\{\xi_k\}$是独立同分布随机变量序列, ξ_k有分布列

$$\begin{pmatrix} 0 & 1 \\ 0.995 & 0.005 \end{pmatrix}, \quad p=P\left(\sum_{k=0}^{10000}\xi_k\leqslant 70\right).$$

利用推论 4，

$$np=10000\times 0.005=50, \quad \sqrt{np(1-p)}\doteq 7.05.$$

$$p=P\left(0\leqslant \sum_{k=1}^{10000}\xi_k\leqslant 70\right)$$

$$=P\left(\frac{-50}{7.05}\leqslant\frac{\sum_{k=1}^{1000}\xi_k-50}{7.05}\leqslant\frac{20}{7.05}\right)$$

$$\doteq\int_{-7.09}^{2.84}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\doteq 0.9977.$$

定理 6 给出了独立随机变量序列$\{\xi_k\}$同分布时的中心极限定理．对于分布不一定相同的情况有

定理 7（李亚普诺夫定理） $\{\xi_k\}$是独立随机变量序列，$E|\xi_k-E\xi_k|^3$存在，$k=1, 2, \cdots$，令

$$B_n=\sqrt{\sum_{k=1}^{n}D\xi_k},$$

若

$$\lim_{n\to+\infty}\frac{1}{B_n^3}\sum_{k=1}^{n}E|\xi_k-E\xi_k|^3=0,$$

则对任何实数 x 有

$$\lim_{n\to+\infty}P\left(\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{B_n}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

习 题 一

1．用依概率收敛的定义证明：若

$$\xi_n\xrightarrow{P}a, \quad \eta_n\xrightarrow{P}b,$$

则 $$\xi_n+\eta_n\xrightarrow{P}a+b, \quad \xi_n\eta_n\xrightarrow{P}ab.$$

2．设 ξ_n 服从参数为 n 的泊松分布，$n=1, 2\cdots$，证明：当 $n\to+\infty$ 时，

$\frac{\xi_n - n}{\sqrt{n}}$ 的分布收敛于 $N(0, 1)$ 的分布.

3. 将 n 个带有号码 1 至 n 的球投入 n 个编有号码 1 至 n 的匣子，并限制每个匣子只能进一个球，设球与匣子号码一致的个数是 S_n，试证

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

[提示：用车贝谢夫不等式.]

4. 设 $\{\xi_n\}$ 独立，同分布，有分布列

$$P(\xi_n = 2^{k-2\ln k}) = 2^{-k} \quad (k=1, 2, \cdots),$$

则 $\{\xi_n\}$ 服从大数定律.

5. 设随机变量序列 $\{\xi_n\}$ 独立，ξ_n 在 $[-n, n]$ 上均匀分布，试问对 $\{\xi_n\}$ 能否用中心极限定理.

[提示：验证李亚普诺夫定理条件.]

第四章小结

这一章介绍了大数定律及中心极限定理.

定理 3，定理 5 使概率论中的两个基本关系：频率与概率，平均值与数学期望——得到了明确的表述.

独立同分布条件下的大数定律和中心极限定理为大样本统计推断(见第五章)提供了理论根据.

从这两类极限定理的证明中可以进一步看到特征函数在研究分布律中的重要作用.

本章中定理的结论都是重要的，至于定理的证明相对说来是次要的. 由于用到的数学工具较深，初学者或非数学专业的读者可以略去.

第五章

统计推断初步

第一节 基本概念

1.1 统计推断概述

在这一章将介绍一些数理统计的基本知识和一些常用的数理统计方法.

在前几章概率论部分,我们引进了一系列的概念,如随机事件的概率,随机变量的分布律及数字特征等等,用这些概念可以在不同程度上刻划我们希望认识的随机现象. 数理统计的任务之一是根据实际观测到的随机试验的结果,对有关事件的概率或随机变量的分布律、数字特征做出估计或推测.

例如,在一定条件下生产出的一件产品可能是合格品,也可能是不合格品. 对这个随机现象可以用"生产一件不合格品"这个随机事件的概率 p(叫废品率)来刻划. 或者引用一个随机变量 ξ, ξ 表示生产出一件产品中的不合格品数, 分布列为

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \tag{1}$$

用随机变量 ξ 刻划这个随机现象. 对于一个生产者来说,知道用废品率 p 或 ξ 的分布列来刻划产品的质量情况,这当然是很重要的,但是,在一定意义上讲,更重要的问题是用什么方法得知 p 的大小或 ξ 的分布列. 在不能或不便于对每件产品进行检验的情况下,通常使用的办法是抽验一部分产品,然

后根据这一部分产品的质量情况对 p 的大小做一个推测或估计. 也有这样的情况，譬如按照某种规定，当 p 不超过 0.01 时，这个生产过程就算是正常的，允许继续生产. 对这种情况就需要根据抽验的一部分产品的质量情况对 p 是否超过 0.01 做一个判断. 因为只有两种可能：$p\leqslant 0.01$ 或 $p>0.01$，所以对这个问题也可以这样叙述：需要根据抽验的一部分产品的质量情况对“$p\leqslant 0.01$”(或“$p>0.01$)做一个判断.

所谓抽验一部分产品，其实就是让 ξ 重复取若干个值.被抽验的第一件产品是合格的就是 ξ 取“0”，被抽验的第二件产品是不合格的就是 ξ 取“1”…. 抽验了 20 件产品的结果就是 ξ 取了 20 个值，如 0, 1, 0, 0, …, 1, 0. 采用这样的看法，上面提出的了解产品质量的问题可以这样叙述：ξ 有分布列(1)，其中 p 是未知的. 现在要根据 ξ 取的若干个(有限个)值 $x_1, x_2, \cdots, x_n$，

(I) 对 p 的大小做一个估计；

(II) 对一个陈述句：“$p\leqslant p_0$”做出判断，肯定它还是否定它？接受它还是拒绝它？p_0 是预先指定的数.

在数理统计中称问题(I)为参数估计问题，称问题(II)为假设检验问题. 这是统计推断中讨论的两个主要问题. 对这两个问题更一般的提法是：ξ 是一个随机变量，它的分布函数 $F(x;\theta)$(或概率分布密度 $p(x;\theta)$，或分布列 $p(x_i;\theta)$, $i=1, 2, \cdots$)的函数形式是已知的，但 θ 是未知参数(实数或有限维实向量). θ 有一个取值范围 Θ，叫参数空间. 如分布列(1)中，p 是未知参数，$\Theta=[0,1]$是参数空间. 现在对 ξ 进行 n 次观察得到 n 个观察值 $x_1, x_2, \cdots, x_n$. 根据 ξ 的观察值对 $F(x;\theta)$ 中的未知参数 θ 的大小做出估计的问题叫参数估计问题. 根据 ξ 的观察值对关于 θ 的一个陈述句“$\theta\in\Theta_0$”(Θ_0

是 Θ 的一个子集)作出接受或拒绝的判断问题叫对参数的假设检验问题. 称陈述句 H: "$\theta\in\Theta_0$" 为统计假设. 习惯上也称为"零假设",相反的陈述句 K: "$\theta\in\Theta-\Theta_0$" 称为对立假设或备择假设. 问题(II)中"$p\leqslant p_0$" 就是一个统计假设,那里的 $\Theta_0=[0, p_0]$. 也有这样的情况, $F(x;\theta)$ 的函数形式也是未知的, 我们希望判断一下它是不是和某一个已知分布函数 $F_0(x)$ 一样. 例如判断一下刻划某一随机现象的随机变量 ξ 是不是参数为 (a_0, σ_0^2) 的正态随机变量. 这时可以把"$F(x;\theta)=F_0(x)$" 做为统计假设,然后根据 ξ 的样本做出接受或拒绝的判断. 称这样的问题为对分布的假设检验问题. 参数估计问题和假设检验问题统称统计推断问题.

1.2 样本和统计量

进行统计推断需要利用随机变量 ξ 的 n 次取值. 观察随机变量 ξ 重复取的 n 个值可以看作一个随机试验. 记第 1 次观察结果为 X_1, 第 n 次观察结果为 X_n, 则 $X_1, \cdots, X_n$ 是 n 个随机变量. 如果这 n 次观察是在相同的条件下独立地重复进行的, 那么, $X_1, \cdots, X_n$ 相互独立, 并且每个 X_i $(i=1, \cdots, n)$ 和 ξ 有相同的分布. 统计推断中使用的观察值通常都是在相同条件下独立地重复进行的.

定义 称随机变量 $X_1, X_2, \cdots, X_n$ 为随机变量 ξ 的一个样本, 如果 $X_1, X_2, \cdots, X_n$ 独立, 并且每个 $X_i(i=1, 2, \cdots, n)$ 与 ξ 有相同的分布,称 n 为样本的容量(或大小).

在前面提到的 ξ 取定的 n 个值 $x_1, x_2, \cdots, x_n$ 就是随机样本 $X_1, X_2, \cdots, X_n$ 取定的值. 事实上 $(x_1, x_2, \cdots, x_n)$ 是对 ξ 进行 n 次重复观察这一随机试验的一个样本点. 今后称 $x_1, x_2, \cdots, x_n$ 为样本的观察值. 在样本的容量 n 给定后,

一个样本观察值是 n 个有序的数，对应着 n 维空间中的一个点(即样本点)，因而样本空间是 n 维空间的一个子集．样本和样本的观察值之间的区别是：前者为随机变量，后者为取定的数．不过为了叙述上的方便，今后有时把样本的观察值也叫做样本，在符号使用上也不去区分它们．至于什么时候它表示样本，什么时候它表示样本的观察值，读者不难从上下文分析清楚．另外在数理统计教科书中，在讲样本之前一般先定义一个"具有分布 $F(x)$ 的总体(或母体)"，实际上可以把这个总体理解为我们这里所说的随机变量 ξ.

参数估计问题和假设检验问题，都是利用样本对未知分布或已知分布中的未知参数进行统计推断．为什么样本能够起这个作用呢？这是因为样本包含有(或能提供)关于未知分布或未知参数的信息．仍以抽验产品为例．分布列(1)中含有未知参数 p, ξ 的样本 $x_1, x_2, \cdots, x_n$ 是由一些"0"和"1"组成的一串数．常识告诉我们，如果废品率 p 比较大，那么样本中的"1"一般会多一些，相对地说"0"会少一些；反之，如果 p 比较小，那么样本中的"1"一般会少一些，"0"会多一些．样本中"1"的多少一般反映着未知参数 p 的大小，这说明样本中包含有关于未知参数 p 的信息．更确切一点说，ξ 的随机样本 $X_1, X_2, \cdots, X_n$ 的分布

$$P(X_1=x_1, \cdots, X_n=x_n)=p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n-\sum\limits_{i=1}^{n} x_i},$$
$$x_i=0 \text{ 或 } 1 \quad (i=1, 2, \cdots, n)$$

与未知参数 p 有关系．样本 $X_1, X_2, \cdots, X_n$ 取各种值的可能性大小受 p 的制约，反过来，$X_1, X_2, \cdots, X_n$ 取的值就提供关于 p 的信息．当然，样本能提供给我们的信息有些和未知参数有关，是需要利用的，有些则和未知参数无关，可以抛弃．通

常用样本的一个或若干个函数来代替样本，目的是简化数据并把有用的信息集中地充分地反映出来. 例如上述x_1, x_2, …, x_n不仅提供了“1”的个数，还提供了“1”的位置. 前者与p有关，当用$\sum_{i=1}^{n} x_i$代替样本时，虽然$\sum_{i=1}^{n} x_i$不再提供关于“1”的位置的信息，但关于“1”的个数的信息在$\sum_{i=1}^{n} x_i$中保存了，集中了.

定义 设$f(x_1, \cdots, x_n)$是定义在样本空间上，不依赖于未知参数的一个函数(实函数或实向量函数)，则随机变量(向量)$f(X_1, \cdots, X_n)$称为一个统计量.

例如随机变量$\xi \sim N(a, \sigma^2)$, a未知, σ^2已知, X_1, X_2, …, X_n是ξ的样本，则

$$f_1(X_1, \cdots, X_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$f_2(X_1, \cdots, X_n) = X_1 + 1,$$

$$f_3(X_1, \cdots, X_n) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

都是统计量. 而

$$f_4(X_1, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$$

就不是统计量，因为它依赖于未知参数a.

样本的矩是重要的统计量.

定义 $X_1, X_2, \cdots, X_n$是随机变量ξ的样本. 对任何自然数r，称$\frac{1}{n}\sum_{i=1}^{n} X_i^r$为样本的$r$阶矩，记作$a_r$. 特别记$a_1$为$\overline{X}$. 称$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$为样本的$r$阶中心矩，记作$m_r$. 特别记$m_2$为$s^2$.

1.3 一些统计量的分布

在实际问题中，用正态随机变量刻划的随机现象是比较普遍的，因此正态样本统计量在数理统计中占有重要的地位．下面我们讨论几个重要的正态样本统计量的分布．

定理 1 $X_i \sim N(0, 1)$, $i=1, 2, \cdots, n$ 独立，则

$$ns^2=\sum_{i=1}^{n}(X_i-\overline{X})^2$$

是自由度为 $n-1$ 的 χ^2 变量，$\overline{X}\sim N\left(0, \frac{1}{n}\right)$，并且 ns^2 与 $\overline{X}$ 独立．

【证】 由线性代数知，存在第一行为$\left(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)$的 n 阶正交方阵 $C=(c_{ij})$．令

$$Y_1=\frac{1}{\sqrt{n}}(X_1+\cdots+X_n),$$
$$Y_i=c_{i1}X_1+\cdots+c_{in}X_n, \quad i=2, 3, \cdots, n,$$

则 $\sum\limits_{i=1}^{n}X_i^2=\sum\limits_{i=1}^{n}Y_i^2$, $Y_1=\sqrt{n}\ \overline{X}$．由引理，$Y_1, \cdots, Y_n$ 独立，每一个都是标准正态随机变量．又

$$ns^2=\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$$
$$=\sum_{i=1}^{n}Y_i^2-Y_1^2=\sum_{i=2}^{n}Y_i^2$$

是 $n-1$ 个独立标准正态随机变量的平方和，所以 $ns^2 \sim\chi^2(n-1)$． $\overline{X}=\frac{1}{\sqrt{n}}Y_1$，所以 $\overline{X}\sim N\left(0, \frac{1}{n}\right)$．因为 ns^2 只是 $Y_2, \cdots, Y_n$ 的函数．$\overline{X}=\frac{1}{\sqrt{n}}Y_1$ 只是 Y_1 的函数，所以 ns^2 与 $\overline{X}$ 独立． ▌

推论1 $X_i\sim N(a,\sigma^2)$, $i=1,2,\cdots,n$ 独立,则

$$\frac{ns^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1),$$

$\overline{X}\sim N\left(a,\frac{\sigma^2}{n}\right)$, 并且 $\frac{ns^2}{\sigma^2}$ 与 $\overline{X}$ 独立.

【证】 令 $Y_i=\frac{X_i-a}{\sigma}$, 则 $Y_i\sim N(0,1)$, $i=1,2,\cdots,n$, $\overline{Y}=\frac{\overline{X}-a}{\sigma}$. 由定理1

$$\frac{ns^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-a}{\sigma}-\frac{\overline{X}-a}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}(Y_i-\overline{Y})^2\sim\chi^2(n-1)$$

与 $\overline{X}=\sigma\overline{Y}+a$ 独立. $\overline{X}\sim N\left(a,\frac{\sigma^2}{n}\right)$是显然的. ┐

推论2 $X_{11},\cdots,X_{1n_1}$; $X_{21},\cdots,X_{2n_2}$; $\cdots$; $X_{p1},\cdots,X_{pn_p}$ 独立, $X_{ij}\sim N(a_i,\sigma^2)$, $j=1,2,\cdots,n_i$, $i=1,2,\cdots,p$. 令

$$\overline{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij},$$

$$s_i^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2,$$

$$i=1,2,\cdots,p,$$

则 $\overline{X}_1,\cdots,\overline{X}_p$, $n_1s_1^2,\cdots,n_ps_p^2$ 独立.

【证】 由于 $\overline{X}_i$, $n s_i^2$ 只是 $X_{i1},X_{i2},\cdots,X_{in_i}$ 的函数, $(\overline{X}_1,n_1s_1^2)$, $(\overline{X}_2,n_2s_2^2)$, $\cdots$, $(\overline{X}_p,n_ps_p^2)$独立. 由推论1及独立随机变量的定义, 对任意给定的 $2p$ 个实数 z_{11}, z_{12}; z_{21}, z_{22}; $\cdots$; z_{p1}, z_{p2}. 有

$$P(\overline{X}_1<z_{11},\ n_1s_1^2<z_{12},\ \overline{X}_2<z_{21},\ n_2s_2^2<z_{22},\ \cdots,$$
$$\overline{X}_p<z_{p1},\ n_ps_p^2<z_{p2})$$
$$=P(\overline{X}_1<z_{11},\ n_1s_1^2<z_{12})P(\overline{X}_2<z_{21},\ n_2s_2^2<z_{22})\cdots$$
$$\times P(\overline{X}_p<z_{p1},\ n_ps_p^2<z_{p2})$$
$$=P(\overline{X}_1<z_{11})P(n_1s_1^2<z_{12})P(\overline{X}_2<z_{21})P(n_2s_2^2<z_{22})\cdots$$
$$\times P(\overline{X}_p<z_{p1})P(n_ps_p^2<z_{p2}).$$ ▌

定理2 $X_i\sim N(a,\ \sigma^2)$, $i=1,\ 2,\ \cdots,\ n$, 独立,则

$$t=\frac{\overline{X}-a}{s}\sqrt{n-1}$$

是自由度为 $n-1$ 的 t 变量，其中 $s=\sqrt{s^2}$.

【证】 由推论 1

$$\overline{X}\sim N\left(a,\ \frac{\sigma^2}{n}\right),\ (\overline{X}-a)\Big/\frac{\sigma}{\sqrt{n}}\sim N(0,\ 1).$$

由推论 1 知，$ns^2/\sigma^2\sim\chi^2(n-1)$ 与 $(\overline{X}-a)\Big/\frac{\sigma}{\sqrt{n}}$ 独立,根据 t 变量的定义,

$$\frac{\overline{X}-a}{s}\sqrt{n-1}=\frac{\overline{X}-a}{\sigma/\sqrt{n}}\Big/\sqrt{\frac{ns^2}{\sigma^2}\Big/(n-1)}$$

$\sim t(n-1)$. ▌

定理3 $X_1,\ X_2,\ \cdots,\ X_m;\ Y_1,\ Y_2,\ \cdots,\ Y_n$ 独立，$X_i\sim N(a,\sigma^2)$, $1\leqslant i\leqslant m$, $Y_j\sim N(b,\ \sigma^2)$, $1\leqslant j\leqslant n$. 令

$$\overline{X}=\frac{1}{m}\sum_{i=1}^{m}x_i,\quad \overline{Y}=\frac{1}{n}\sum_{j=1}^{n}Y_j,$$

$$s_x^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\overline{X})^2,\quad s_y^2=\frac{1}{n}\sum_{j=1}^{n}(Y_j-\overline{Y})^2,$$

则

(i)
$$\frac{(\overline{X}-\overline{Y})-(a-b)}{\sqrt{\dfrac{ms_x^2+ns_y^2}{m+n-2}\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}}$$

是自由度为 $m+n-2$ 的 t 变量;

(ii) $$\frac{n-1}{m-1}\cdot\frac{ms_x^2}{ns_y^2}$$

是第一自由度为 $m-1$, 第二自由度为 $n-1$ 的 F 变量.

【证】 (i) 由推论 2, $\overline{X}$, $\overline{Y}$, ms_x^2, ns_y^2 独立. 又

$$\overline{X}\sim N\left(a,\ \frac{\sigma^2}{m}\right),\ \overline{Y}\sim N\left(b,\ \frac{\sigma^2}{n}\right),$$

$$\frac{ms_x^2}{\sigma^2}\sim\chi^2(m-1),\quad \frac{ns_y^2}{\sigma^2}\sim\chi^2(n-1).$$

所以

$$\frac{(\overline{X}-\overline{Y})-(a-b)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim N(0,\ 1),$$

$$\frac{ms_x^2+ns_y^2}{\sigma^2}\sim\chi^2(m+n-2)$$

而且独立. 根据 t 变量的定义

$$\frac{(\overline{X}-\overline{Y})-(a-b)}{\sqrt{\frac{ms_x^2+ns_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}=\frac{(\overline{X}-\overline{Y})-(a-b)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}\Bigg/\sqrt{\frac{ms_x^2+ns_y^2}{\sigma^2}\Big/(m+n-2)}$$

$\sim t(m+n-2)$.

(ii) 根据 F 变量的定义

$$\frac{n-1}{m-1}\cdot\frac{ms_x^2}{ns_y^2}=\frac{ms_x^2}{\sigma^2(m-1)}\Bigg/\frac{ns_y^2}{\sigma^2(n-1)}$$

$\sim F(m-1,\ n-1)$. 】

习 题 一

1. 利用公式

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\Big/n,\quad \bar{x}=\sum_{i=1}^{n}x_i\Big/n,$$

计算样本

$$0.31,\ 0.26,\ 0.28,\ 0.33,\ 0.29,\ 0.32,\ 0.24,\ 0.25$$

的二阶中心矩.

2. 令 $y_i=x_i-c$, $i=1, 2, \cdots, n$, $\bar{y}=\left(\sum_{i=1}^{n} y_i\right)\Big/n$, 则有

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

从第1题8个数据中都减去0.28得

$$0.03,\ -0.02,\ 0,\ 0.05,\ 0.01,\ 0.04,\ -0.05,\ -0.04;$$

试计算二阶中心矩.

3. 令 $z_i=cx_i$, $i=1, 2, \cdots, n$, $\bar{z}=\sum_{i=1}^{n} z_i\Big/n$, 则有

$$\sum_{i=1}^{n}(z_i-\bar{z})^2=c^2\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

将第2题中的8个数据都乘以100, 再用上面等式计算二阶中心矩.

4. $X_1, \cdots, X_n$; $Y_1, \cdots, Y_n$ 独立, $X_i\sim N(a, \sigma_1^2)$, $Y_j\sim N(b, \sigma_2^2)$, $1\leqslant i, j\leqslant n$. 令

$$s_{xy}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}),$$

试证

$$\frac{(\bar{X}-\bar{Y})-(a-b)}{\sqrt{s_x^2+s_y^2-2s_{xy}}}\sqrt{n-1}$$

服从自由度为 $n-1$ 的 t 分布.

5. $\xi\sim N(a, \sigma^2)$, $X_1, X_2, \cdots, X_n$ 是 ξ 的样本, 证明

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^2=(n^2-1)\sigma^4.$$

第二节　未知参数的估计

2.1 点估计

设 $F(x;\theta)$ 是随机变量 ξ 的分布函数, θ 是未知参数. 首先考虑 θ 是实数的情况. 所谓对 θ 的点估计就是给一个规则, 按照这个规则, 只要得到一个样本的观察值 $x_1, x_2, \cdots, x_n$, 就

能够给 θ 一个确定的估计值. 例如 $\theta=E\xi$, 用样本的平均值 $\bar{x}$ 作为 θ 的估计值就是给了一个规则, 按照这个规则, 当我们得到样本

$$1.2,\ 1.0,\ 1.2,\ 0.8,\ 0.9,\ 1.1,\ 1.0$$

时, 就用

$$\bar{x}=\frac{1}{7}(1.2+1.0+1.2+0.8+0.9+1.1+1.0)\doteq 1.03$$

作为 θ 的估计值. 给一个规则, 实际上就是给了一个定义在样本空间上的函数. 这个函数当然不能依赖于未知参数 θ, 所以这个函数就是在第一节定义的统计量. 为了强调将用这样的函数对未知参数作估计, 有时称统计量为估计量. 用来对 θ 作估计的估计量记作 $\hat{\theta}(X_1, \cdots, X_n)$, 对一个取定的样本的观察值 $x_1, \cdots, x_n$, 对应的 $\hat{\theta}(x_1, \cdots, x_n)$ 叫 θ 的估计值.

推广到未知参数 $\theta=(\theta_1, \cdots, \theta_r)$ 是 r 维实向量的情况, 或更一般地考虑被估计的参数是 θ 的函数

$$g(\theta)=(g_1(\theta_1, \cdots, \theta_r), \cdots, g_s(\theta_1, \cdots, \theta_r))$$

的情况, 所谓对 $g(\theta)$ 的点估计就是确定 s 个统计量

$$\hat{g}_i(X_1, \cdots, X_n), \quad i=1, \cdots, s,$$

$$\hat{g}(X_1, \cdots, X_n)=(\hat{g}_1(X_1, \cdots, X_n), \cdots, \hat{g}_s(X_1, \cdots, X_n))$$

叫 $g(\theta)$ 的估计量. 当我们得到样本 $x_1, \cdots, x_n$ 后, 就用对应的 $\hat{g}(x_1, \cdots, x_n)$ 作为 $g(\theta)$ 的估计值.

1. 确定估计量的方法

这里介绍两个常用的方法.

(1) 矩法

设 $X_1, X_2, \cdots, X_n$ 是随机变量 ξ 的样本. 用矩法确定估计量, 是指用样本的 r 阶矩 $a_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r$ 作为未知参数 $E\xi^r$

的估计量,用样本矩的函数 $h(a_1, \cdots, a_m)$ 作为相应的未知参数 $h(E\xi, \cdots, E\xi^m)$ 的估计量. 例如用矩法确定的 $D\xi=E\xi^2-(E\xi)^2$ 的估计量是 $a_2-a_1^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$. 由大数定律知 $n\to+\infty$ 时, $a_r\xrightarrow{P}E\xi^r$, 如果 $h(t_1, \cdots, t_m)$ 是连续函数时, 还有 $h(a_1, \cdots, a_m)\xrightarrow{P}h(E\xi, \cdots, E\xi^m)$. 所以当样本容量比较大时,用矩法确定的估计量一般依概率收敛于被估计参数.

(2) 极大似然法

先结合下面的例子来说明.

设 $\xi\sim N(\theta, 1)$, $\theta\in(-\infty, +\infty)$ 是未知参数, $X_1, \cdots, X_n$ 是 ξ 的样本,它的联合概率分布密度是

$$L(x_1, x_2, \cdots, x_n; \theta)=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}. \qquad (2)$$

对任一组 $x_1, x_2, \cdots, x_n$, 视(2)为 θ 的函数,求出使 L 极大的 θ 值. 容易看出, 当 $\theta=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 时,

$$L(x_1, x_2, \cdots, x_n; \theta)=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta)^2\right]}$$

取极大值. 我们就用 $\overline{X}$ 作为 θ 的估计量.

用 $\overline{X}$ 作为 θ 的估计量是这样想的: 对指定的 θ, 视 $L(x_1, \cdots, x_n; \theta)$ 为 $x_1, \cdots, x_n$ 的函数, 可以表示随机样本取 $x_1, \cdots, x_n$ 值的相对可能性大小. 当然, 对不同的 θ, 取 $x_1, \cdots, x_n$ 值的相对可能性大小有所不同, 有的大一些, 有的小一些. 但是,基于上述想法,当我们已经得到样本的观察值 $x_1, \cdots, x_n$, 要回答 θ 是什么值时, 我们有理由认为是使取 $x_1, \cdots, x_n$ 的相对可能性最大的 θ. 当然, 这是一个朴素的想法, 可以证明用这种方法确定的估计量在一定条件下有一些

好的性质，例如它依概率收敛于被估计的参数。

一般，若 ξ 是连续型随机变量，有概率分布密度 $p(x;\theta)$，$\theta\in\Theta$，样本 $X_1, \cdots, X_n$ 的联合概率分布密度

$$L(x;\theta)=L(x_1, \cdots, x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta), \tag{3}$$

视为 θ 的函数，称为样本的似然函数。

若 ξ 是离散型随机变量，有分布列

$$p(x;\theta),\ x=a_i,\ i=1, 2, \cdots;\ \theta\in\Theta.$$

样本 $X_1, X_2, \cdots, X_n$ 的分布列

$$L(x;\theta)=L(x_1, \cdots, x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta),$$

$$x_i=a_{l_i},\ l_i=1, 2, \cdots,\ i=1, 2, \cdots, n, \tag{4}$$

视为 θ 的函数，称为样本的似然函数。

若存在函数 $\hat{\theta}(x_1, \cdots, x_n)$ 使

$$L(x_1, \cdots, x_n;\hat{\theta}(x_1, \cdots, x_n))=\sup_{\theta\in\Omega}L(x_1, \cdots, x_n;\theta)$$

对几乎一切 $(x_1, x, \cdots, x_n)$ 成立，即有一 N 使上式对一切 $(x_1, \cdots, x_n)\in\overline{N}$ 成立，其中 N 满足 $P_\theta[(X_1, \cdots, X_n)\in N]=0$，$\theta\in\Theta$，则称 $\hat{\theta}(X_1, \cdots, X_n)$ 为 θ 的极大似然估计量。

【例 1】 $\xi\sim N(a, \sigma^2)$，$\theta=(a, \sigma^2)\in(-\infty, +\infty)\times(0, +\infty)=\Theta$ 是未知参数，试求 θ 的极大似然估计量。

解：ξ 的样本 $X_1, X_2, \cdots, X_n$ 的似然函数是

$$L(x;a, \sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right\},$$

视 $L(x;a, \sigma^2)$ 为 (a, σ^2) 的函数。因为 $\ln L(x;a, \sigma^2)$ 与 $L(x;a, \sigma^2)$ 在相同的点上取得极值，所以可以求使 $\ln L(x;a, \sigma^2)$ 极大的 (a, σ^2) 值。

$$\ln L(x;\ a,\ \sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2$$

$$=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$-\frac{1}{2\sigma^2}n(\bar{x}-a)^2.$$

$$\begin{cases}\dfrac{\partial}{\partial a}\ln L(x;\ a,\ \sigma^2)=\dfrac{n}{\sigma^2}(\bar{x}-a)=0,\\[2mm] \dfrac{\partial}{\partial \sigma^2}\ln L(x;\ a,\ \sigma^2)=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2=0.\end{cases}\tag{5}$$

解方程组(5)得

$$a=\bar{x},\quad \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=s^2.\tag{6}$$

因为

$$\ln L(x;\ \bar{x},\ s^2)-\ln L(x;\ a,\ \sigma^2)$$

$$=\left[-\frac{n}{2}\ln s^2-\frac{n}{2}\right]-\left[-\frac{n}{2}\ln(\sigma^2)-\frac{ns^2}{2\sigma^2}\right.$$

$$\left.-\frac{n}{2\sigma^2}(\bar{x}-a)^2\right]$$

$$=-\frac{n}{2}\ln\frac{s^2}{\sigma^2}-\frac{n}{2}+\frac{ns^2}{2\sigma^2}+\frac{n}{2\sigma^2}(\bar{x}-a)^2$$

$$\geqslant\frac{n}{2}\left[\left(\frac{s^2}{\sigma^2}-1\right)-\ln\frac{s^2}{\sigma^2}\right]\geqslant 0,$$

所以

$$\ln L(x;\ \bar{x},\ s^2)=\sup_{(a,\sigma^2)\in\Theta}\ln L(x;\ a,\ \sigma^2),$$

$\hat{\theta}(x)=(\overline{X},\ s^2)$是$\theta=(a,\ \sigma^2)$的极大似然估计量.

【例2】 ξ在$[0,\ \theta]$上均匀分布，$\theta>0$是未知参数，试求θ的极大似然估计量.

解：样本 $X_1, X_2, \cdots, X_n$ 的似然函数是

$$L(x_1, \cdots, x_n; \theta)=\begin{cases}\dfrac{1}{\theta^n}, & \text{当 } 0<\min\limits_{1\leqslant i\leqslant n}\{x_i\},\ \max\limits_{1\leqslant i\leqslant n}\{x_i\}\leqslant\theta; \\ 0, & \text{其它情况.}\end{cases} \tag{7}$$

取 $\overline{N}=(0, +\infty)\times\cdots\times(0, +\infty)$，则有 $P_\theta[(X_1, \cdots, X_n)\in N]=0$，$0<\theta<+\infty$，对任何 $(x_1, x_2, \cdots, x_n)\in\overline{N}$，视 $L(x_1, \cdots, x_n; \theta)$ 为 θ 的函数，由(7)知，θ 在 $\max\limits_{1\leqslant i\leqslant n}\{x_i\}$ 处取极大值．所以

$$\hat{\theta}(X_1, \cdots, X_n)=\max_{1\leqslant i\leqslant n}\{X_i\}$$

是 θ 的极大似然估计量．

【例 3】 ξ 有概率分布密度

$$p(x;\ \theta)=\frac{1}{2}e^{-|x-\theta|},\ -\infty<x<+\infty,$$

$\theta\in(+\infty, +\infty)$ 是未知参数，试求 θ 的极大似然估计量．

解：ξ 的样本 $X_1, \cdots, X_n$ 的似然函数是

$$L(x;\ \theta)=\frac{1}{2^n}e^{-\sum\limits_{i=1}^{n}|x_i-\theta|},$$

$$\ln L(x;\ \theta)=-n\ln 2-\sum_{i=1}^{n}|x_i-\theta|,$$

任何使 $\sum\limits_{i=1}^{n}|x_i-\theta|$ 极小的 θ 能使 $L(x;\ \theta)$ 极大．将 $x_1, x_2, \cdots, x_n$ 从小到大排成 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$．当 $n=2k+1$ 时，$x_{(k+1)}$ 使 $\sum\limits_{i=1}^{n}|x_i-\theta|$ 极小，$x_{(k+1)}$ 是 θ 的极大似然估计量．当 $n=2k$ 时，$[x_{(k)}, x_{(k+1)}]$ 上的任何值都使 $\sum\limits_{i=1}^{n}|x_i-\theta|$ 极小．所以，$X_{(k+1)}$ 是 θ 的一个极大似然估计量．

【例 4】 从某一个湖里捉出 1000 尾鱼，分别涂上红点后

放回．隔了一段时间以后，再捉出1000尾鱼，发现其中有100尾涂有红点．试根据这些情况估计湖里鱼的总数n．

解：用ξ表示第二次捉出的1000尾鱼中涂有红点鱼的个数，则ξ有分布列

$$p_k(n)=\frac{C_{1000}^{k}C_{n-1000}^{1000-k}}{C_{n}^{1000}},\quad k=0,\ 1,\ 2,\ \cdots,\ 1000.$$

现在ξ已经取了一个容量为1的样本$x_1=100$，我们将根据这个样本求n的极大似然估计．

由于

$$\begin{aligned}\frac{p_k(m)}{p_k(m-1)}&=\frac{C_{1000}^{k}C_{m-1000}^{1000-k}}{C_{m}^{1000}}\Bigg/\frac{C_{1000}^{k}C_{m-1-1000}^{1000-k}}{C_{m-1}^{1000}}\\&=\frac{C_{m-1000}^{1000-k}\cdot C_{m-1}^{1000}}{C_{m-1001}^{1000-k}\cdot C_{m}^{1000}}=\frac{(m-1000)(m-1000)}{(m-2000+k)m}\\&=\frac{(m-2000)m+1000000}{(m-2000)m+mk},\end{aligned}$$

当$mk\leqslant 1000000$时，$p_k(m)/p_k(m-1)\geqslant 1$，$p_k(m)$随着$m$增大而不减少；当$mk>1000000$时，$p_k(m)/p_k(m-1)<1$，$p_k(m)$随着$m$增大而减小，所以当$m$为不超过$1000000/k$的最大整数时，$p_k(m)$有最大值．本例$k=100$，所以$n$的极大似然估计值为10000．

2．鉴定估计量好坏的准则

前面我们给出了两个确定估计量的方法．但是从估计量只是样本空间到参数空间的一个对应规则来看，对未知参数的估计量是没有太多限制的．你可以这样估计(采用这样的对应规则)，他可以那样估计(采用那样的对应规则)．例如，为了估计$\theta=E\xi$，可以用估计量

$$\hat{\theta}_1(X_1,\ \cdots,\ X_n)=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

$$\hat{\theta}_2(X_1, \cdots, X_n)=(\max_{1\le i\le n}\{X_i\}+\min_{1\le i\le n}\{X_i\})/2,$$

也可以用 $\hat{\theta}_3(X_1, \cdots, X_n)\equiv 0.$

在这些估计量中，人们会觉得采用 $\hat{\theta}_1$ 或 $\hat{\theta}_2$ 比采用 $\hat{\theta}_3$ 好一点，因为用 $\hat{\theta}_3$ 作为 θ 的估计量就是不管 ξ 的数学期望 θ 是什么，都用 0 作为 θ 的估计值，而丝毫不考虑样本提供的关于 θ 的信息. 但是用 $\hat{\theta}_1$ 或 $\hat{\theta}_2$ 作为 θ 的估计量又好在什么地方呢？对这个问题可能也很难一下子回答清楚. 实际上人们在使用某一估计量时，脑子里会有这样的或那样的，明确的或不十分明确的"好"的标准. 这里有必要列出一些常用的标准，以便用它对各种估计量的优缺点作出鉴定和选择.

(1) 无偏性

称估计量 $\hat{g}(X_1, \cdots, X_n)$是参数 $g(\theta)$的无偏估计量，如果对任何 $\theta\in\Theta$ 有 $E_\theta\hat{g}(X_1, \cdots, X_n)=g(\theta)$或 $E_\theta[\hat{g}(X_1, \cdots, X_n)-g(\theta)]=0$. 这里 $E_\theta\hat{g}$ 表示 $\hat{g}$ 的分布律中的参数是 θ 时，$\hat{g}$ 的数学期望. 为了书写简便，今后有时也用 $E\hat{g}$ 代替 $E_\theta\hat{g}$.

对这个准则可以这样理解. 设 $g(\theta)$ 是一维参数，估计量 $\hat{g}(X_1, \cdots, X_n)$是一个随机变量，而 $g(\theta)$是一个确定的数（虽然我们不知道它的值），当得到样本的观察值 $x_1, \cdots, x_n$ 后，用 $\hat{g}(x_1, \cdots, x_n)$作为 $g(\theta)$的估计值时，在一般情况下会有一个偏差 $\hat{g}(x_1, \cdots, x_n)-g(\theta)$（虽然我们不知道它是多少）. 这个偏差可能是正的，可能是负的. 象这样对 $g(\theta)$ 的一次估计中出现一个偏差是不足为奇的. 但是如果用估计量 $\hat{g}$ 多次对 $g(\theta)$ 进行估计，而偏差的平均值不为 0，例如大于 0，这就表明总的来看，或做为一种估计方法，用 $\hat{g}$ 估计 $g(\theta)$ 是平均偏高的；相反的，如果偏差的平均值小于 0，则表明做为一种估计方法，用 $\hat{g}$ 估计 $g(\theta)$ 是平均偏低的. 要求估计量是无偏的，

就是要求它作为一种估计方法既不偏高，也不偏低.

【例 5】 ξ 有分布函数 $F(x;\theta)$，未知参数 θ 是 ξ 的数学期望. 试问用样本 $X_1, \cdots, X_n$ 的平均值 $\overline{X}$ 作为 θ 的估计量是不是无偏估计量?

解：由样本的定义，X_i 和 ξ 有相同的分布，因而

$$EX_i = E\xi = \theta, \quad i=1, 2, \cdots, n;$$

又

$$E_\theta \overline{X} = E_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E_\theta X_i = \frac{1}{n}\cdot n\theta = \theta.$$

所以 $\overline{X}$ 是 θ 的无偏估计量.

【例 6】 ξ 有分布函数 $F(x;a,\sigma^2)$，未知参数 a, σ^2 分别是 ξ 的数学期望与方差. 试问用样本的二阶中心矩

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

作为 σ^2 的估计量，是不是无偏估计量?

解：$X_1, X_2, \cdots, X_n$ 独立，每一个 X_i 与 ξ 有相同的分布，因而

$$EX_i = a, \quad DX_i = \sigma^2, \quad i=1, 2, \cdots, n.$$

又

$$\begin{aligned}
&E_{(a,\sigma^2)} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= \frac{1}{n} E_{(a,\sigma^2)} \sum_{i=1}^{n}[(X_i - a) - (\overline{X} - a)]^2 \\
&= \frac{1}{n} E_{(a,\sigma^2)} \left[\sum_{i=1}^{n}(X_i - a)^2 - n(\overline{X} - a)^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} E(X_i - a)^2 - nE(\overline{X} - a)^2\right] \\
&= \frac{1}{n}(n\sigma^2 - nD\overline{X}) = \sigma^2 - D\overline{X}.
\end{aligned}$$

所以 $D\overline{X}=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$,

$$Es^2=E\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2.$$

用 s^2 作为 σ^2 的估计量不是无偏估计量．但在上式两端乘以 $\frac{n}{n-1}$ 可以得到

$$E\frac{n}{n-1}s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\sigma^2,$$

这说明 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是 σ^2 的无偏估计量，许多数理统计书称 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 为样本的方差．

【例 7】 $\xi\sim N(a,\sigma^2)$，a、σ^2 是未知参数．上例证明了 $\frac{n}{n-1}s^2$ 是 σ^2 的无偏估计量，试问

$$\sqrt{\frac{n}{n-1}s^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

是不是 σ 的无偏估计量？

解：$\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$ 是自由度为 $n-1$ 的 χ^2 变量，

$$E\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}}=\int_0^{+\infty}x\frac{2}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}x^{n-2}e^{-\frac{x^2}{2}}dx$$

$$=\frac{\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^{+\infty}\left(\frac{x^2}{2}\right)^{\frac{n}{2}-1}e^{-\frac{x^2}{2}}d\left(\frac{x^2}{2}\right)$$

$$=\frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

$$\therefore\quad E\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}=\sqrt{\frac{2}{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\sigma$$

$$\neq\sigma,$$

$\sqrt{\frac{n}{n-1}s^2}$ 不是 σ 的无偏估计量.

(2) 有效性

一个未知参数 θ 的无偏估计量有时不只一个. 在例 5 中, 如果 $n>1$, 显然 x_1 就是 θ 的另一个无偏估计量. 更一般的, 任意选 n 个实数 $c_1, c_2, \cdots, c_n$, 只要 $c_1+c_2+\cdots+c_n=1$, $\sum_{i=1}^{n}c_iX_i$ 都是 θ 的无偏估计量. 如果 $D\xi$ 存在, 我们还可以比较这些无偏估计量的方差以进一步比较这些估计量的优劣. 例如

$$D\overline{X}=\frac{1}{n}D\xi<D\xi=DX_1,$$

说明用 $\overline{X}$ 和 X_1 估计 θ 虽然都既不偏高也不偏低; 但用 $\overline{X}$ 作为 θ 的估计量得到的估计值, 总的看更接近于 θ. 从这个意义上讲, 我们愿意用 $\overline{X}$ 而不愿意用 X_1 作为 θ 的估计量, 而且称 $\overline{X}$ 比 X_1 有效.

一般的, 设 $\hat{g}_1(X_1, \cdots, X_n)$, $\hat{g}_2(X_1, \cdots, X_n)$ 都是 $g(\theta)$ 的无偏估计量, 如果

$$D\hat{g}_1\leqslant D\hat{g}_2,$$

就说 $\hat{g}_1$ 比 $\hat{g}_2$ 有效.

【例 8】 ξ 有分布函数 $F(x;\theta)$, $\theta=E\xi$ 是未知参数, $D\xi$ 存在, $X_1, X_2, \cdots, X_n$ 是 ξ 的样本. 考虑一切形如 $\sum_{i=1}^{n}c_iX_i$ 的估计量, 其中 $c_1, c_2, \cdots, c_n$ 是任何 n 个满足 $\sum_{i=1}^{n}c_i=1$ 的实数, 则由

$$E\left(\sum_{i=1}^{n} c_i X_i\right)=\sum_{i=1}^{n} c_i(EX_i)=\sum_{i=1}^{n} c_i\theta=\theta\left(\sum_{i=1}^{n} c_i\right)=\theta$$

知，$\sum_{i=1}^{n} c_i X_i$ 是 θ 的无偏估计量．又

$$D\left(\sum_{i=1}^{n} c_i X_i\right)=\left(\sum_{i=1}^{n} c_i^2\right) D\xi \geqslant \frac{1}{n}\left(\sum_{i=1}^{n} c_i\right)^2 D\xi$$

$$=\frac{1}{n} D\xi=D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right),$$

所以 $\overline{X}$ 比一切形如 $\sum_{i=1}^{n} c_i X_i\left(\sum_{i=1}^{n} c_i=1\right)$ 的估计量都有效．

(3) 最小均方误差

在例题 4 中曾经看到，以无偏性作为鉴别估计量好坏的标准，用 $\frac{n}{n-1}s^2$ 作为 $D\xi$ 的估计量比 s^2 好．但是，如果用另外的标准衡量，也不一定．在选定一个参数 $g(\theta)$ 的估计量 $\hat{g}(X_1, \cdots, X_n)$ 后，在一般情况下，估计值与真正的 $g(\theta)$ 有一个偏差 $|\hat{g}(x_1, \cdots, x_n)-g(\theta)|$．人们希望估计的平均偏差能比较小，因此可以用均方误差 $E[\hat{g}(X_1, \cdots, X_n)-g(\theta)]^2$ 的大小作为衡量估计量好坏的标准．

【例 9】 $\xi\sim N(a, \sigma^2)$，用 $\frac{1}{c}\sum_{i=1}^{n}(X_i-\overline{X})^2$，$(n>1)$ 作为 σ^2 的估计量，

$$E\left[\frac{1}{c}\sum_{i=1}^{n}(X_i-\overline{X})^2-\sigma^2\right]^2$$

$$=E\left[\frac{1}{c^2}\left\{\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}^2-\frac{2\sigma^2}{c}\sum_{i=1}^{n}(X_i-\overline{X})^2+\sigma^4\right]$$

$$=\frac{1}{c^2}E\left\{\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}^2-\frac{2\sigma^2}{c}E\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]+\sigma^4$$

$$=\frac{1}{c^2}(n^2-1)\sigma^4-\frac{2(n-1)}{c}\sigma^4+\sigma^4$$

是 $\frac{1}{c}$ 的二次多项式, 在 $\frac{2(n-1)}{2(n^2-1)}=\frac{1}{n+1}$ 处有极小值, 故

$$
\begin{aligned}
&E\left[\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\overline{X})^2-\sigma^2\right]^2 \\
&<E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2-\sigma^2\right]^2 \\
&<E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2-\sigma^2\right]^2.
\end{aligned}
$$

这个例题说明: 如果用 $E[\hat{g}(X_1,\cdots,X_n)-g(\theta)]^2$ 的大小衡量估计量的好坏, 有时 $D\xi$ 的有偏估计 s^2 比无偏估计 $\frac{n}{n-1}s^2$ 好. 当然, 这两个估计量都不如 $\frac{n}{n+1}s^2$ 好.

(4) 一致性

当样本容量 $n\to+\infty$ 时, 如果估计量 $\hat{g}(X_1,\cdots,X_n)$ 依概率收敛(几乎处处收敛)于被估计参数 $g(\theta)$, 就称 $\hat{g}(X_1,\cdots,X_n)$ 是 $g(\theta)$ 的弱(强)一致估计量.

用一致性衡量估计量, 一个好的估计量应当随着样本容量的增大能以更大的概率取得接近于被估计参数的值. 这是一个很自然的要求. 用矩法得到的许多估计量都具有这种性质.

2.2 区间估计

在 2.1 中我们讲了对未知参数的点估计问题. 我们能用一个统计量(是随机变量)作为某一个未知参数(常量)的点估计量就是因为这个统计量对被估计的参数来说有一些好的性质. 例如, 我们用随机变量 ξ 的样本平均值 $\overline{X}$ 作为 $E\xi$ 的估计量, 是因为 $\overline{X}$ 对 $E\xi$ 来说有无偏性和一致性等好的性质, 这样, $\overline{X}$ 取值虽然有随机性, 但总的来看, 用它估计 $E\xi$ 既不

偏高，也不偏低；而且，当样本容量比较大时，$\overline{X}$ 取的值多数靠近 $E\xi$，当实际问题中提出要了解 $E\xi$ 等于什么的时候，我们就用 $\bar{x}$ 做为 $E\xi$ 的估计值了．但是在使用点估计时，有时也感到有不足的地方，例如我们只是大体上知道样本容量"比较大"时，$\overline{X}$ 取的值"多数""靠近" $E\xi$，但是"比较大"大到什么程度，"多数"多到什么程度，"靠近"近到什么程度还是不清楚的．下面要讲的区间估计方法在一定程度上弥补了这个不足．以下用 P_θ 表示参数值为 θ 时的概率．

所谓对未知参数 $g(\theta)$ 做区间估计就是确定两个统计量 $f_1(X_1, X_2, \cdots, X_n) < f_2(X_1, X_2, \cdots, X_n)$，使

$$P_\theta[f_1(X_1, \cdots, X_n) < g(\theta) < f_2(X_1, \cdots, X_n)] = \beta, \quad (8)$$

$0<\beta<1$ 是预先给定的数，叫置信系数，当我们得到样本值 $x_1, x_2, \cdots, x_n$ 时，我们就估计 $g(\theta)$ 在下面区间内

$$(f_1(x_1, \cdots, x_n), f_2(x_1, \cdots, x_n)). \quad (9)$$

【例 10】 $\xi \sim N(a, 1)$，想用样本 $X_1, \cdots, X_n$ 对未知参数 a 做区间估计．由于 $\overline{X} \sim N\left(a, \dfrac{1}{n}\right)$，在指定 β，例如 $\beta = 0.95$ 后，就有

$$P_a\left(\frac{|\overline{X}-a|}{1/\sqrt{n}} < 1.96\right) = 0.95$$

或 $$P_a\left(\overline{X} - 1.96\frac{1}{\sqrt{n}} < a < \overline{X} + 1.96\frac{1}{\sqrt{n}}\right) = 0.95.$$

在这个例中 $$f_1(X_1, \cdots, X_n) = \overline{X} - \frac{1.96}{\sqrt{n}},$$

$$f_2(X_1, \cdots, X_n) = \overline{X} + \frac{1.96}{\sqrt{n}}. \quad (10)$$

如果我们得到的样本是

5.1, 5.1, 4.8, 5.0, 4.7, 5.0, 5.2, 5.1, 5.0,

代入(10)得

$$f_1(x_1,\ \cdots,\ x_n)=5-\frac{1.96}{\sqrt{9}}\doteq 4.35,$$

$$f_2(x_1,\ \cdots,\ x_n)=5+\frac{1.96}{\sqrt{9}}\doteq 5.65.$$

我们就估计 a 在区间(4.35, 5.65)内．

区间估计所回答的是未知参数在什么范围内，而不是未知参数等于什么。

例10中估计 a 在区间(4.35, 5.65)内对吗？为了回答这个问题，我们回过头来理解一下(8)式．

在(8)中，$g(\theta)$ 虽是未知的，但是是一个确定的数．样本 $X_1,\ \cdots,\ X_n$ 是随机变量，因而 $f_1(X_1,\ \cdots,\ X_n)$，$f_2(X_1,\ \cdots,\ X_n)$ 也是随机变量，以 f_1、f_2 为左右端点可以构成一个随机开区间．一旦得到样本值，区间 $(f_1,\ f_2)$ 就定下来了．这时，或者 $g(\theta)\in(f_1,\ f_2)$，或者 $g(\theta)\overline{\in}(f_1,\ f_2)$，更确切地说，区间 $(f_1,\ f_2)$ 或者套上 $g(\theta)$，或者套不上 $g(\theta)$，因而

$$f_1(X_1,\ \cdots,\ X_n)<g(\theta)<f_2(X_1,\ \cdots,\ X_n),$$

即随机区间 $(f_1,\ f_2)$ 套上 $g(\theta)$ 是一个随机事件．(8)式表示这个事件的概率是 β，至于这个事件在一次具体试验(取一个样本的观察值)中是否出现，我们就不知道了．回到刚才提出的问题，估计 a 在区间(4.35, 5.65)内对不对，决定于区间(4.35, 5.65)是否套上了 a，这在一次具体的估计中是无法知道的．但是，有一点我们是有把握的，这就是估计 a 在 $\bar{x}-1.96\frac{1}{\sqrt{n}}$ 和 $\bar{x}+1.96\frac{1}{\sqrt{n}}$ 之间成功的概率是0.95．如果你使用这个方法对 a 估计100次，那么大体上有95次左右估计是正确的，虽然你并不知道哪一次是正确的．特别，用置信系数接近于1的区间估计方法，估计的结果几乎就是正确的。

【例 11】 $\xi \sim N(a,\ \sigma^2)$, a, σ^2 未知，试用样本 $X_1,\ \cdots,\ X_n$ 分别对 a, σ^2 做区间估计，置信系数取 $1-\alpha$.

从例 10 知，最重要的是找到一个类似于 $\sqrt{n}(\overline{X}-a)$ 的仅仅是样本及未知参数的函数，而且它的分布是已知的.

为了对 a 做区间估计，我们选用函数

$$t=\frac{\overline{X}-a}{s}\sqrt{n-1},$$

由第一节定理 2 知，它服从自由度为 $n-1$ 的 t 分布，故可以从 t 分布表中查得 t_α 使

$$P_a\left(\frac{|\overline{X}-a|}{s}\sqrt{n-1}<t_\alpha\right)=1-\alpha;$$

$$P_a\left(\overline{X}-t_\alpha\frac{s}{\sqrt{n-1}}<a<\overline{X}+t_\alpha\frac{s}{\sqrt{n-1}}\right)=1-\alpha.$$

我们就用 $\left(\overline{X}-t_\alpha\frac{s}{\sqrt{n-1}},\ \overline{X}+t_\alpha\frac{s}{\sqrt{n-1}}\right)$ 作为 a 的区间估计.

为了对 σ^2 做区间估计，我们选用函数

$$\chi^2=\frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\frac{ns^2}{\sigma^2},$$

由第一节推论 1 知，它服从自由度为 $n-1$ 的 χ^2 分布，故可以从 χ^2 分布表中查得 b_1, b_2 使

$$P(\chi^2_{n-1}\leqslant b_1)=\frac{\alpha}{2},\ P(\chi^2_{n-1}\geqslant b_2)=\frac{\alpha}{2};$$

因而
$$P\left(b_1<\frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}<b_2\right)=1-\alpha,$$

$$P\left(\frac{ns^2}{b_2}<\sigma^2<\frac{ns^2}{b_1}\right)=1-\alpha.$$

我们就用 $\left(\frac{ns^2}{b_2},\ \frac{ns^2}{b_1}\right)$ 作为 σ^2 的区间估计.

如某糖厂用自动打包机装糖，每包重量ξ服从正态分布$N(a, \sigma^2)$. 某日开工后测得9包重量如下：

$$99.3,\ 98.7,\ 100.5,\ 101.2,\ 98.3,$$
$$99.7,\ 99.5,\ 102.1,\ 100.5,$$

现在使用这9包糖的重量对a做区间估计，置信系数用0.95. 查t分布表得$t_\alpha=2.31$.

$$\bar{x}-t_\alpha\frac{s}{\sqrt{n-1}}=99.978-2.31\times\frac{1.143}{\sqrt{8}}=99.04,$$

$$\bar{x}+t_\alpha\frac{s}{\sqrt{n-1}}=99.978+2.31\times\frac{1.143}{\sqrt{8}}=100.91.$$

结论：估计a在区间(99.04, 100.91)内.

【例 12】 $\xi\sim N(a, \sigma^2)$, $\eta\sim N(b, \sigma^2)$ 有相同的方差σ^2, 试用ξ的样本$X_1, \cdots, X_m$, η的样本$Y_1, \cdots, Y_n$(两个样本是独立的)对$b-a$做区间估计.

解：由于$X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ 独立, $\overline{X}\sim N\left(a, \frac{\sigma^2}{m}\right)$, $\overline{Y}\sim N\left(b, \frac{\sigma^2}{n}\right)$是独立的,

$$\overline{Y}-\overline{X}\sim N\left(b-a, \frac{\sigma^2}{m}+\frac{\sigma^2}{n}\right).$$

若σ^2已知，利用函数

$$\frac{(\overline{Y}-\overline{X})-(b-a)}{\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}}\sim N(0, 1),$$

可以对$b-a$做出区间估计.

若σ^2未知，则由第一节定理3知

$$t=\frac{(\overline{Y}-\overline{X})-(b-a)}{\sqrt{\frac{ms_x^2+ns_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

服从自由度为$m+n-2$的t分布，给定置信系数$1-\alpha$后，从

t 分布表查 t_α 使

$$P(|t|\geqslant t_\alpha)=\alpha,$$

我们就用

$$\left(\overline{Y}-\overline{X}-t_\alpha\sqrt{\frac{ms_x^2+ns_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)},\right.$$
$$\left.\overline{Y}-\overline{X}+t_\alpha\sqrt{\frac{ms_x^2+ns_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}\right)$$

做为 $b-a$ 的区间估计.

在用区间估计方法时，有一个成功概率大小的问题及估计范围大小的问题. 人们在解决实际问题时总是希望估计的范围小一点，成功的概率大一点. 但是在样本容量一定的情况下二者是不能兼得的.

在例 10 中，如果希望估计成功的概率再大一点，例如希望 $\beta=0.997$，由于

$$P_a\left(\frac{|\overline{X}-a|}{1/\sqrt{n}}<3\right)=0.997,$$

可以用 $\left(\overline{X}-3\frac{1}{\sqrt{n}},\ \overline{X}+3\frac{1}{\sqrt{n}}\right)$ 作为 a 的区间估计，但这时估计区间的长度为 $6\frac{1}{\sqrt{n}}$，比原来的 $3.92\frac{1}{\sqrt{n}}$ 大了. 如果希望缩小估计区间的长度，例如取 $\left(\overline{X}-\frac{1}{\sqrt{n}},\ \overline{X}+\frac{1}{\sqrt{n}}\right)$ 为 a 的区间估计，这时估计区间长度为 $\frac{2}{\sqrt{n}}$，短了，但估计成功的概率只有

$$P\left(\frac{|\overline{X}-a|}{1/\sqrt{n}}<1\right)=0.683,$$

比 0.95 小多了. 如果希望估计成功的概率为 0.997，又希望估计区间的长度为 $\frac{2}{\sqrt{n}}$，那只好把样本的容量 n 扩大成原来的 9 倍，这时

$$P_a\left(\frac{|\overline{X}-a|}{1/\sqrt{9n}}<3\right)=0.997,$$

$$P_a\left(\overline{X}-\frac{1}{\sqrt{n}}<a<\overline{X}+\frac{1}{\sqrt{n}}\right)=0.997.$$

当然，在实际问题中取容量太大的样本有时也是不容易的，甚至是做不到的，因此在解决实际问题时需要权衡各方面的情况，确定合适的 n 和 β.

习 题 二

1. ξ 在 $[0, \theta]$ 上均匀分布，$0<\theta<+\infty$，试用 ξ 的样本 $X_1, X_2, \cdots, X_n$ 求 θ 的矩法估计.
2. ξ 在 $[\theta_1, \theta_2]$ 上均匀分布，$-\infty<\theta_1<\theta_2<+\infty$，试用 ξ 的样本 $X_1, X_2, \cdots, X_n$ 求 θ_1, θ_2 的矩法估计.
3. ξ 服从参数为 λ 的泊松分布，试用 ξ 的样本 $X_1, X_2, \cdots, X_n$ 求 λ 的极大似然估计.
4. ξ 在 $\left[\theta-\frac{1}{2}, \theta+\frac{1}{2}\right]$ 上均匀分布，$-\infty<\theta<+\infty$，试用 ξ 的样本 $X_1, X_2, \cdots, X_n$ 求 θ 的极大似然估计.
5. ξ 有概率分布密度

$$p(x;\theta)=\begin{cases}\theta x^{\theta-1}, & 当\ 0<x<1;\\ 0, & 其它,\end{cases}$$

$0<\theta<+\infty$，试用样本 $X_1, X_2, \cdots, X_n$ 求 θ 的极大似然估计.

6. ξ 在 $[0, \theta]$ 上均匀分布，$0<\theta<+\infty$，试用 ξ 的样本 $X_1, X_2, \cdots, X_n$ 求 $E\xi$、$D\xi$ 的极大似然估计.
7. ξ 在 $[0, \theta]$ 上均匀分布，$X_1, \cdots, X_n$ 是 ξ 的样本，证明

$$\frac{n+1}{n}\max\{X_1, \cdots, X_n\},\quad (n+1)\min\{X_1, \cdots, X_n\}$$

都是 θ 的无偏估计，并比较哪一个更有效.

8. ξ 在 $[\theta-a, \theta+a]$ 上均匀分布，$a>0$ 已知，$X_1, X_2, \cdots, X_n$ 是 ξ 的样本. 证明 $\frac{1}{n}\sum_{i=1}^{n}X_i$，$\frac{1}{2}(\max\limits_{1\le i\le n}\{X_i\}+\min\limits_{1\le i\le n}\{X_i\})$ 都是 θ 的无偏

估计，并比较那一个更有效。

9. 设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是参数 θ 的两个独立的无偏估计，$D\hat{\theta}_1=kD\hat{\theta}_2>0$，$k$ 已知．试找出常数 k_1，k_2 使 $k_1\hat{\theta}_1+k_2\hat{\theta}_2$ 是 θ 的无偏估计，并且在所有这样的线性估计中方差最小．

10. $\xi\sim N(a,\sigma^2)$，$\sigma^2>0$ 未知．甲得到 ξ 的样本 $X_1, X_2, \cdots, X_{n_1}$，并用 $\hat{\sigma}_1^2=\frac{1}{n_1-1}\sum\limits_{i=1}^{n_1}(X_i-\overline{X})^2$ 估计 σ^2；乙得到 ξ 的样本 $Y_1, Y_2, \cdots, Y_{n_2}$，并用 $\hat{\sigma}_2^2=\frac{1}{n_2-1}\sum\limits_{i=1}^{n_2}(Y_i-\overline{Y})^2$ 估计 σ^2．如果两个样本是独立的，试用 $\hat{\sigma}_1^2$，$\hat{\sigma}_2^2$ 构造优于 $\hat{\sigma}_1^2$，$\hat{\sigma}_2^2$ 的 σ^2 的无偏估计．

11. $\xi\sim N(a,\sigma_1^2)$，$\eta\sim N(b,\sigma_2^2)$ 独立，σ_1^2，σ_2^2 未知，试利用第一节习题中第 4 题构造对未知参数 $a-b$ 的区间估计方法．若 ξ 的样本为

$$5.1,\ 5.2,\ 5.7,\ 5.3,\ 5.3,$$

η 的样本为

$$2.1,\ 2.5,\ 2.6,\ 2.6,\ 2.2,$$

试求 $a-b$ 的区间估计．置信系数取 0.95．

第三节　假设检验

在第一节中已经谈到，假设检验是根据随机变量 ξ 的样本 $x_1, \cdots, x_n$ 对关于 ξ 分布的一个统计假设 H 做出接受或拒绝的判断．这和估计未知参数一样，需要给出一个规则，当我们得到一个样本后，根据这个规则就能做出接受 H 或拒绝 H 的决定．所谓给出一个规则实际上就是确定样本空间到只包含两个元的集合{拒绝，接受}上的一个对应．在这个对应下，如果样本空间中的点 $(x_1', \cdots, x_n')$ 对应着"拒绝"，那么，当得到样本 $(x_1', \cdots, x_n')$ 时，就拒绝 H；如果样本空间中的点 $(x_1'', \cdots, x_n'')$ 对应着"接受"，那么，当得到样本 $(x_1'', \cdots, x_n'')$ 时，就接受 H．因为样本空间中任何一点，不对应着"拒绝"就对应着"接受"，所以，只要把对应于"拒绝"的点明确了，一个对应规则或一个检验方法也就确定了．一切对应于"拒绝"的点

组成样本空间的一个区域，我们将称这个区域为判别区域（或否定域、临界域）.

现在让你根据 ξ 的样本 $x_1, \cdots, x_n$ 对一个关于 ξ 分布的统计假设 H 作检验，如果把判别区域 S 也告诉了你，这个检验你就会做了. 当取了一个样本点 $(x_1, \cdots, x_n)$ 之后，若 $(x_1, \cdots, x_n)\in S$ 就拒绝 H，否则就接受 H. 例如 $\xi\sim N(a, 1)$，a 是未知参数，希望用容量为 9 的样本 $x_1, \cdots, x_9$ 检验统计假设 H：$a=1.5$，判别区域取

$$S=\{(x_1, \cdots, x_9)\,|\,|\bar{x}-1.5|\geqslant 0.6\}$$

（选择这样一个判别区域的想法是 $\bar{x}$ 与 1.5 差别太大时应当拒绝 H），若取得样本

$$x_0=(1.6,\ 2.0,\ 2.1,\ 1.7,\ 1.8,\ 1.8,\ 1.7,\ 2.0,\ 1.5),$$

由于 $\bar{x}=1.8$，$x_0\bar{\in} S$，按照约定的规则应当接受 H. 需要进一步考虑的问题是怎样选取判别区域 S. 能不能在样本空间中随便选一个区域做判别区域呢？这和确定未知参数 $g(\theta)$ 的估计量时遇到的问题一样，就是用这个判别区域确定的检验规则或检验方法好不好？接着的问题是什么叫好？怎样选出一个好的判别区域，即怎样确定一个好的检验方法呢？为了回答这些问题，我们先看一看按照某一个判别区域对统计假设 H 做检验时，检验的效果是什么.

3.1 假设检验中的两类错误

在第一节检验产品的例中，ξ 表示生产一件产品中的废品数，它的分布列中含有未知参数 p，参数空间为闭区间 $[0, 1]$，我们关心的是“$p\leqslant 0.01$”对不对. 为了检验它，可以提出统计假设

$$H:\ p\leqslant 0.01.$$

当我们拒绝 H 时，就要接受 $p>0.01$. 把 H 作为零假设，$p>0.01$ 就是备择假设, 用 K 表示. 所谓提出一个统计假设检验问题就是提出零假设和备择假设. 现在取一个容量为 100 的样本 $x_1, \cdots, x_{100}$, 也就是抽验 100 件产品, 得到

$$x_i=\begin{cases}0, & \text{当抽验的第 } i \text{ 件产品不是废品;}\\ 1, & \text{当抽验的第 } i \text{ 件产品是废品,}\end{cases} \quad i=1, 2, \cdots, 100.$$

当这 100 件产品中废品比较多时，自然要怀疑零假设的正确性，所以可以指定一个数，例如就是 2，当废品数 $\sum_{i=1}^{100} x_i>2$ 时就拒绝零假设，否则就接受零假设，这时判别区域就是

$$S=\left\{(x_1, \cdots, x_{100}) \,\middle|\, \sum_{i=1}^{100} x_i>2\right\}.$$

现在重申一下，在给了判别区域后，对 H 是拒绝还是接受，就看样本是否属于 S. 例如，得到一个样本，算出 $\sum_{i=1}^{100} x_i=3$，由于这个样本属于 S，我们的结论是拒绝零假设. 在拒绝零假设，即做出废品率超过 0.01 这样的判断后，就应当采取措施降低废品率. 不过人们可以提出疑问：废品率真的超过了 0.01 吗？只根据抽验的 100 件产品中废品数大于 2 就做出 $p>0.01$ 的判断不会错吗？这个疑问是有道理的. 可能有这样的情况，零假设 H 真，即 $p\leqslant 0.01$（虽然我们不知道），但因为取样是随机的，在抽验的 100 件产品中刚好碰上较多的废品，例如有 3 个，按照检验规则（我们只能按照这个规则）应当拒绝 H，这样我们就做了一个错误的判断——把真的零假设拒绝了. 象这样的错误能够避免吗？要想避免这样的错误，除非对任何样本都接受零假设（即 S 是空集）. 即使 $p>0.01$（虽然我们不知道）也接受零假设，认为 $p\leqslant 0.01$. 这就导致另一个错误的判断，把假的零假设接受了. 我们将称拒绝真

的零假设的错误为第一类错误；称接受假的零假设的错误为第二类错误．为了避免犯第一类错误就不可避免地要犯第二类错误．有没有可能选择一个判别区域使两类错误都不出现呢？回答是：只要用抽样检验的方法，不论怎样选择判别区域，总是要犯错误的．粗略地说，当样本的容量给定后，判别区域越"小"，犯第一类错误的可能性越小，而犯第二类错误的可能性越"大"；反之，判别区域越"大"，则犯第一类错误的可能性越"大"，犯第二类错误的可能性越"小"．

一般的，设随机变量 ξ 有分布函数 $F(x;\ \theta)$，θ 是未知参数，参数空间为 Θ．关于参数 θ 的假设检验问题的一般形式是：

$$H:\ \theta\in\Theta_0;\quad K:\ \theta\in\Theta-\Theta_0.$$

$\Theta_0\subseteq\Theta$ 是 Θ 的一个真子集．在确定判别区域 S 及得到样本点 $x=(x_1,\ \cdots,\ x_n)$ 后，可能出现下面四种情况：

(1) $\theta\in\Theta_0$，$x\overline{\in}S$；接受 H．

(2) $\theta\in\Theta_0$，$x\in S$；拒绝 H．

(3) $\theta\in\Theta-\Theta_0$，$x\overline{\in}S$；接受 H．

(4) $\theta\in\Theta-\Theta_0$，$x\in S$；拒绝 H．

(1)、(4)两种情况，判断与实际一致；(2)、(3)则不一致．情况(2)是 H 真，但遭到拒绝，出现第一类错误；情况(3)是 H 不真，但被接受，出现第二类错误．第一类错误只在 H 真时才有可能出现，而出现与否取决于 x 是否属于 S，所以

$$P_\theta\,(\text{犯第一类错误})=\begin{cases}P_\theta(X\in S), & \theta\in\Theta_0;\\ 0, & \theta\in\Theta-\Theta_0.\end{cases}$$

第二类错误只在 H 不真时才有可能出现，出现与否取决于 x 是否不属于 S，所以

$$P_\theta\,(\text{犯第二类错误})=\begin{cases}0, & \theta\in\Theta_0;\\ P_\theta(X\overline{\in}S), & \theta\in\Theta-\Theta_0.\end{cases}$$

特别，当 ξ 有概率分布密度 $p(u;\ \theta)$ 时，

$$P_\theta\,(\text{犯第一类错误})=\begin{cases}\int\cdots\int\limits_S(p(x_1;\ \theta)\cdots p(x_n;\ \theta)dx_1-dx_n, \\ \qquad\qquad \theta\in\Theta_0,\\ 0, \quad \theta\in\Theta-\Theta_0;\end{cases}$$

P_θ (犯第二类错误)

$$=\begin{cases}0, & \theta\in\Theta_0,\\ 1-\int\cdots\int\limits_S p(x_1;\ \theta)\cdots p(x_n;\ \theta)dx_1\cdots dx_n, & \theta\in\Theta-\Theta_0.\end{cases}$$

当 ξ 是离散型随机变量时有类似的表达式。

前面已经分析过，在用样本 x 对 H 做检验时，无论 S 怎样选，两类错误都想避免是不可能的；我们只能希望选的 S 使犯两类错误的可能性小一点。通常采用的选择 S 的原则是使犯第一类错误的概率不要超过给定的数 $\alpha(0<\alpha<1)$，同时，使犯第二类错误的概率尽可能地小。α 通常取 0.01, 0.05 或 0.10，叫作检验的信度或显著性水平。怎样根据这个原则选择 S 将在 3.2 介绍，在此之前，先定义几个概念。

定义 称由判别区域 S 确定的检验方法为水平为 $\alpha(0<\alpha<1)$ 的检验方法，如果

$$P_\theta(X\in S)\leqslant\alpha,\quad \theta\in\Theta_0.$$

从定义知，水平为 α 的检验方法指犯第一类错误的概率不超过 α 的检验方法。今后把“检验方法”简称为“检验”，称由判别区域 S 确定的检验方法为“检验 S”。

定义 称水平为 α 的检验 S 在 $\theta_0\in\Theta-\Theta_0$ 上最强，如果对任何水平为 α 的检验 S' 有

$$P_{\theta_0}(x\in S)\geqslant P_{\theta_0}(x\in S').$$

从定义知，在一切水平为 α 的检验中，检验 S 在 $\theta=\theta_0$ 时不犯第二类错误的概率最大.

定义 称水平为 α 的检验 S 在 $\Theta-\Theta_0$ 上一致最强，如果 S 在 $\Theta-\Theta_0$ 的每一点上最强.

3.2 一致最强检验

考虑一个简单的情况. 设 $\xi\sim N(\theta, 1)$，未知参数 θ 的空间 $\Theta=\{1, 2\}$ 只包含两个点. 假设检验问题是 $H: \theta=1$; $K: \theta=2$. 现在用容量为 1 的样本 X 对 H 做检验，样本空间就是实直线. 给定 $\alpha=0.05$，我们希望解决的问题是在实直线上选择一个区域 S，使得

（i）$\theta=1$ 时，犯第一类错误，即 $X\in S$ 的概率

$$P_1(X\in S)\leqslant\alpha;$$

（ii）$\theta=2$ 时，犯第二类错误，即 $X\bar{\in} S$ 的概率

$$P_2(X\bar{\in} S)=1-P_2(X\in S)$$

最小，换句话说是使 $P_2(X\in S)$ 最大.

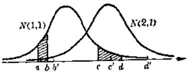


图 5-1

图 5-1 给出了 $N(1, 1)$ 和 $N(2, 1)$ 的图象，把它们分别叫作曲线(I)、(II)吧. 如果要 S 只满足 (1)，那是容易做到的. 任何一个区间，例如图 5-1 中的区间 $[a, b]$，它和曲线(I)围成的曲边梯形的面积是 0.04，因为

$$P_1(X\in[a, b])=0.04<0.05,$$

所以 $[a, b]$ 就是满足(1)的一个 S. 区间 $[c, d]$ 也是一个满足(i)的 S. 任何一个区间，只要它和曲线(I)围成的曲边梯形面积不超过 0.05，都是满足(i)的 S. 但是，如果再加上条件

(ii)，情况就不同了，要在满足条件(i)的S中挑一个使$P_2(X\in S)$(如果S是一个区间，$P_2(X\in S)$就是S和曲线(II)围成的曲边梯形的面积)最大的S. 区间$[a, b]$显然不是这样的S，因为把$[a, b]$放大为$[a, b']$使$P_1(X\in[a, b'])=0.05$，$P_2(X\in[a, b'])$就比$P_2(X\in[a, b])$大，所以我们只需在使$P_1(X\in S)=0.05$的S中选择使$P_2(X\in S)$最大的S. 区间$[a, b']$是不是这样的S呢？从图上直观地看，$P_2(X\in[c, d])$要比$P_2(X\in[a, b'])$大得多. 区间$[c, d]$是不是呢？例如区间$[c', d']$也是一个满足(i)的S，$P_2(X\in[c', d'])$是不是比$P_2(X\in[c, d])$更大呢？这只靠在图上看就不容易看出来了. 不过，有一点是明确的，我们要找的S首先应当包含这样一些点，在这些点上曲线(II)相对说来要比曲线(I)"高"得多. 我们将根据这个想法给出确定满足条件(i)、(ii)的S的方法.

下面的两个定理都是著名的奈门-皮尔逊(Neyman-Pearson)基本引理的特殊情形.

定理1 ξ是连续型随机变量，有分布密度$p(x;\theta)$，$\theta\in\Theta=\{\theta_0, \theta_1\}$. 对假设检验问题

$$H:\ \theta=\theta_0;\quad K:\ \theta=\theta_1$$

及信度α，如果有一个实数k使样本空间中的区域

$$S=\{(x_1, \cdots, x_n)\,|\,p(x_1;\theta_1)\cdots p(x_n;\theta_1)\geqslant kp(x_1;\theta_0)\cdots p(x_n;\theta_0)\}\tag{1}$$

满足

$$P_{\theta_0}((X_1, \cdots, X_n)\in S)=\int\cdots\int_S p(x_1;\theta_0)\cdots p(x_n;\theta_0)dx_1\cdots dx_n=\alpha,\tag{2}$$

则S是水平为α的最强检验.

【证】 由(2)，S显然是水平为α的检验. 又(1)中$k>0$,

否则(2)中 $\alpha=1$ 与 $0<\alpha<1$ 矛盾. 设 S' 是任一水平为 α 的检验,即

$$P_{\theta_0}((X_1, \cdots, X_n)\in S') = \int\cdots\int_{S'} p(x_1;\ \theta_0)\cdots p(x_n;\ \theta_0)dx_1\cdots dx_n \leqslant \alpha,$$

如果能够证明

$$\int\cdots\int_{S} p(x_1;\ \theta_1)\cdots p(x_n;\ \theta_1)dx_1\cdots dx_n \geqslant \int\cdots\int_{S'} p(x_1;\ \theta_1)\cdots p(x_n;\ \theta_1)dx_1\cdots dx_n,$$

S 就是水平为 α 的一致最强检验了.

令 $L(x;\ \theta_i)=p(x_1;\ \theta_i)\cdots p(x_n;\ \theta_i),\quad i=0,\ 1.$

根据 S 的定义,在 $S-S'$ 上, $L(x;\ \theta_1)\geqslant kL(x;\ \theta_0)$; 在 $S'-S$ 上, $L(x;\ \theta_1)<kL(x;\ \theta_0)$, 因而有

$$\begin{aligned}
&\int\cdots\int_{S} L(x;\ \theta_1)dx_1\cdots dx_n-\int\cdots\int_{S'} L(x;\ \theta_1)dx_1\cdots dx_n\\
&=\int\cdots\int_{S-S'} L(x;\ \theta_1)dx_1\cdots dx_n-\int\cdots\int_{S'-S} L(x;\ \theta_1)dx_1\cdots dx_n\\
&\geqslant k\int\cdots\int_{S-S'} L(x;\ \theta_0)dx_1\cdots dx_n-k\int\cdots\int_{S'-S} L(x;\ \theta_0)dx_1\cdots dx_n\\
&=k\int\cdots\int_{S} L(x;\ \theta_0)dx_1\cdots dx_n-k\int\cdots\int_{S'} L(x;\ \theta_0)dx_1\cdots dx_n\\
&\geqslant k(\alpha-\alpha)=0. \quad ▌
\end{aligned}$$

定理1告诉我们, 为了给检验问题$(H,\ K)$确定一个最强检验,关键是找到一个合适的 k, 用它确定判别区域(1).

【例1】 $\xi\sim N(a,\ \sigma^2)$, σ^2 已知, 现在用定理1的方法给检验问题

$$H:\ a=a_0;\quad K:\ a=a_1\quad (a_0<a_1)$$

确定一个最强检验,信度为 0.05.

在这个例中

$$L(x;\ a_i)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_i)^2\right\}$$

相当于(1)的 S 是

$$S=\left\{(x_1,\ \cdots,\ x_n)\ \middle|\ (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_1)^2\right\}\right.$$
$$\left.\geqslant k(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_0)^2\right\}\right\}. \tag{3}$$

这里 k 是待定的,我们希望要的 k 必须满足(2),即

$$P_{a_0}((X_1,\ \cdots,\ X_n)\in S)$$
$$=\int\cdots\int_S(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a_0)^2\right\}dx_1\cdots dx_n$$
$$=0.05. \tag{4}$$

从 S 的表达式(3)看, $k=0$ 时, S 是整个 n 维空间, (4) 式积分为 1; 随着 k 连续增大, 积分连续变小, 有可能找到一个合适的 k 使(4)式成立. 但是,按照这个想法去确定 k 还是有困难的. 现在注意由(3)导出的 S 的另一个表达式

$$S=\left\{(x_1,\ \cdots,\ x_n)\ \middle|\ \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(x_i-a_0)^2\right.\right.$$
$$\left.\left.-\sum_{i=1}^{n}(x_i-a_1)^2\right\}\geqslant\ln k\right\}$$
$$=\left\{(x_1,\ \cdots,\ x_n)\ \sum_{i=1}^{n}x_i\geqslant\frac{2\sigma^2\ln k-n(a_0^2-a_1^2)}{2(a_1-a_0)}\right\}$$
$$=\left\{(x_1,\ \cdots,\ x_n)\ \frac{\bar{x}-a_0}{\sigma/\sqrt{n}}\geqslant c\right\}, \tag{5}$$

其中 $\quad c=\left(\frac{2\sigma^2\ln k-n(a_0^2-a_1^2)}{2n(a_1-a_0)}-a_0\right)\Big/\frac{\sigma}{\sqrt{n}}.$

由于在 $a=a_0$ 时,随机变量

$$\frac{\overline{X}-a_0}{\sigma/\sqrt{n}}\sim N(0,\ 1),$$

查正态分布表知，$c=1.645$ 可使(4)成立，即

$$P_{a_0}((X_1, \cdots, X_n)\in S)=P_{a_0}\left(\frac{\bar{X}-a_0}{\sigma/\sqrt{n}}\geqslant 1.645\right)$$
$$=0.05.$$

在得到 c 后，k 和 S 就都不难得到了，对 H 的检验也就可以进行了．但这些都是不必要的，因为当我们得到一个样本点 $(x_1, \cdots, x_n)$ 后，由(5)看它是否属于 S 可以用 $(\bar{x}-a_0)\Big/\frac{\sigma}{\sqrt{n}}$ 是否不小于 1.645 来判别．总之，按照定理 1 得到的检验方法是：在得到样本 $(x_1, \cdots, x_n)$ 后，先计算平均值 $\bar{x}$.

若 $\bar{x}\geqslant a_0+1.645\sigma/\sqrt{n}$，就拒绝 H；

若 $\bar{x}<a_0+1.645\sigma/\sqrt{n}$，就接受 H.

例如，$a_0=1$，$a_1=2$，$\sigma=1$，$n=25$ 时 $a_0+1.645\sigma/\sqrt{n}=1.329$．当得到的样本平均值 $\bar{x}>1.329$ 时，就拒绝 $E\xi=1$，否则就接受 $E\xi=1$，这是一个水平为 0.05 的最强检验．

这里我们再给上述检验方法做一个直观的解释．要根据 ξ 的样本 $x_1, \cdots, x_n$ 判断 $E\xi=a_0$ 对不对，而且 $E\xi$ 只可能是 a_0，a_1 中的一个．由于 $a_0<a_1$，在 $\bar{x}$ 太大时更有理由怀疑 $E\xi=a_0$ 正确，所以 $\bar{x}$ 太大时应拒绝 H，这就容易理解判别区域的形式是 $[\bar{x}\geqslant c]$．至于 c 取什么值将由信度 α 决定．

例 1 中若 $a_0>a_1$，判别区域

$$S=\left\{(x_1, \cdots, x_n)\ \Big|\ \bar{x}\leqslant\frac{2\sigma^2\ln k-n(a_0^2-a_1^2)}{n(a_1-a_0)}\right\}$$
$$=\left\{(x_1, \cdots, x_n)\ \Bigg|\ \frac{\bar{x}-a_0}{\frac{\sigma}{\sqrt{n}}}\leqslant c'\right\}, \tag{6}$$

$c'=-1.645$．按照定理 1 得到的检验方法是：

若 $\bar{x}\leqslant a_0-1.645\frac{\sigma}{\sqrt{n}}$，就拒绝 H；

若 $\bar{x} > a_0 - 1.645\frac{\sigma}{\sqrt{n}}$，就接受 H.

即 $\bar{x}$ 太小时拒绝 H.

当 ξ 是离散型随机变量时，也有类似的定理，证明方法也是类似的；下面只叙述定理的内容而不再重复定理的证明了.

定理 2 ξ 是离散型随机变量，有分布列 $p(a_i;\ \theta)$，$i=1, 2, \cdots$；$\theta \in \Omega = \{\theta_0, \theta_1\}$. 对假设检验问题

$$H:\ \theta=\theta_0; \quad K:\ \theta=\theta_1$$

及信度 α 如果有一个实数 k 使样本空间中的区域

$$S=\{(a_{l_1}, \cdots, a_{l_n}) \mid p(a_{l_1};\ \theta_1)\cdots p(a_{l_n};\ \theta_1) \geqslant kp(a_{l_1};\ \theta_0)\cdots p(a_{l_n};\ \theta_0)\}, \tag{7}$$

满足

$$P_{\theta_0}((X_1, \cdots, X_n)\in S) = \sum_{(a_{l_1}, \cdots, a_{l_n})\in S} p(a_{l_1};\ \theta_0)\cdots p(a_{l_n};\ \theta_0) = \alpha,$$

则 S 是水平为 α 的最强检验.

【例 2】 ξ 是离散型随机变量，有分布列

$$\begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}.$$

试对检验问题 $H:\ \theta=0.5$；$K:\ \theta=0.6$ 求水平为 1/32 的最强检验. 样本容量为 5.

解：ξ 的分布列可以表示为

$$p(a_i;\ \theta)=\theta^{a_i}(1-\theta)^{1-a_i}, \quad a_1=0, \quad a_2=1.$$

相当定理 2 中的区域(7)为

$$S=\{(a_{l_1}, \cdots, a_{l_5}) \mid 0.6^{\sum_{i=1}^{5} a_{l_i}}(1-0.6)^{5-\sum_{i=1}^{5} a_{l_i}} \geqslant k(0.5)^5\}$$
$$=\left\{(a_{l_1}, \cdots, a_{l_5}) \,\middle|\, \sum_{i=1}^{5} a_{l_i} \geqslant \frac{\ln 1.25^5 k}{\ln 1.5}\right\},$$

其中 k 是待定的，必需满足

$$P_{0.5}((X_1, \cdots, X_5) \in S) = P_{0.5}\left(\sum_{i=1}^{5} x_i \geqslant \frac{\ln 1.25^5 k}{\ln 1.5}\right)$$
$$= \frac{1}{32}. \tag{8}$$

令
$$c = \frac{\ln 1.25^5 k}{\ln 1.5},$$

由于 H 真时 $\sum_{i=1}^{5} x_i$ 服从参数为(5, 0.5)的二项分布, 分布列为:

0	1	2	3	4	5
$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

取 c 为(4, 5)中任何数, (8)式都成立. 和例题1一样, 在找到 c 后不必再计算 k 了, 因为在得到样本点$(x_1, \cdots, x_5)$后, 看它是否属于 S 可以用 $\sum_{i=1}^{5} x_i$ 是否大于4来判别. 所以所求的检验是

若 $\sum_{i=1}^{5} x_i > 4$, 就拒绝 H;

若 $\sum_{i=1}^{5} x_i \leqslant 4$, 就接受 H.

(直观地理解是 $\sum_{i=1}^{5} x_i$ 太大时拒绝 H.)

从例题1、2可以看到, 对这种简单的假设检验问题, 利用定理1、2有可能找到具有一定显著水平的最强检验, 要紧的是定理中的条件: "如果有一个实数 k 使…"必须满足. 否则, 如例2中的 α 为0.05, 则取(4, 5)中任何值为 c 确定的 S 有

$$P_{0.5}((X_1, \cdots, X_5) \in S) = \frac{1}{32} < 0.05,$$

减小 c 把 S 扩大一点, 就有

$$P_{0.5}((X_1,\ \cdots,\ X_5)\in S)=\frac{1}{32}+\frac{5}{32}=\frac{6}{32}>0.05,$$

满足定理条件的 k 不存在了. 在理论上为了解决这样的问题考虑了"随机化检验". 在解决实际问题时, 如果碰到这种情况不妨把 α 改变一下,如例 2 就不要用 0.05,而用 $\frac{1}{32}\doteq 0.03$.

以上我们讨论的假设检验问题中零假设 H: $\theta\in\Theta_0$, Θ_0 中只含有一个点, 备择假设 K: $\theta\in\Theta-\Theta_0$, $\Theta-\Theta_0$ 中也只含有一个点的情况. 称只含有一个点的假设为简单假设, 否则为复合假设. 零假设和备择假设都是简单的, 这在实际问题中是很少见的. 某个产品的废品率一般不会只有两个值: 不是 p_0 就是 p_1. 但是,对零假设和备择假设都是简单的情况的讨论有助于解决某些零假设或备择假设是复合假设的假设检验问题.

【例 3】 $\xi\sim N(a,\ \sigma^2)$, σ 已知. 考虑假设检验问题

$$H:\ a=a_0;\quad K:\ a>a_0.$$

这里零假设是简单的,备择假设 K: $a\in(a_0,\ +\infty)$是复合的. 在给定信度 α 及样本容量 n 后, 为了给这个假设检验问题找一个好的检验方法,先任取 $a_1>a_0$, 考虑检验问题

$$H:\ a=a_0;\quad K':\ a=a_1.$$

由例 1 知, 检验问题(H, K')的水平为 α 的最强检验的判别区域为

$$S=\left\{(x_1,\ \cdots,\ x_n)\ \frac{\bar{x}-a_0}{\sigma/\sqrt{n}}\geqslant c\right\},\tag{9}$$

c 由 $P_{a_0}((X_1,\ \cdots,\ X_n)\in S)=\alpha$ 决定. 这就是说: 如果 S' 是检验问题(H, K')的任一水平为 α 的检验,则有

$$P_{a_1}((X_1,\ \cdots,\ X_n)\in S)\geqslant P_{a_1}((X_1,\ \cdots,\ X_n)\in S').\tag{10}$$

注意: S 和 a_1 的选择无关,因而(10)式对一切 $a_1\in(a_0,\ +\infty)$

成立．所以，以 S 为判别区域的检验也是检验问题 (H, K) 的水平为 α 的一致最强检验．

【例 4】 $\xi\sim N(a, \sigma^2)$，σ^2 已知．考虑假设检验问题

$$H:\ a\leqslant a_0,\quad K:\ a>a_0,$$

a_0 是预先指定的数．在给定信度 α 及样本容量 n 后，先看检验问题

$$H':\ a=a_0;\quad K:\ a>a_0.$$

由例 3 知判别区域(9)对应的检验是检验问题 (H', K) 的水平为 α 的一致最强检验．下面证明这个检验也是检验问题 (H, K) 的水平为 α 的一致最强检验．首先对于任何 $a\leqslant a_0$,

$$\begin{aligned}P_a((X_1,\ \cdots,\ X_n)\in S)&=P_a\left(\frac{\overline{X}-a_0}{\sigma/\sqrt{n}}\geqslant c\right)\\&=P_a\left(\frac{\overline{X}-a}{\sigma/\sqrt{n}}\geqslant c+\frac{a_0-a}{\sigma/\sqrt{n}}\right)\\&\leqslant P_a\left(\frac{\overline{X}-a}{\sigma/\sqrt{n}}\geqslant c\right)=\alpha.\end{aligned}$$

这说明 S 对 H 而言是水平为 α 的检验．又设 S' 是水平为 α 的另一个检验，即

$$P_a((X_1,\ \cdots,\ X_n)\in S')\leqslant\alpha,\quad a\leqslant a_0,$$

特别 $$P_{a_0}((X_1,\ \cdots,\ X_n)\in S')\leqslant\alpha.$$

对任何 $a_1>a_0$，由例 3 知(10)成立，所以 S 是检验问题 (H, K) 的水平为 α 的一致最强检验．

【例 5】 已知某厂生产的某种化学纤维纤度 $\xi\sim N(a, 0.04^2)$．某天测得 25 根纤维的纤度，平均值为 1.41，试问 a 不大于 1.40 是否可信？取信度 $\alpha=0.05$．

解：把 $a\leqslant 1.40$ 作为零假设，考虑检验问题

$$H:\ a\leqslant 1.40;\quad K:\ a>1.40.$$

我们已经得到一个容量为 25 的样本点 $(x_1,\ \cdots,\ x_{25})$，算得平

均值 $\bar{x}=1.41$. 由例题 4 知，检验问题(H, K)的最强检验的判别区域为

$$S=\left\{(x_1, \cdots, x_n)\ \middle|\ \frac{\bar{x}-1.40}{0.04/\sqrt{25}}\geqslant c\right\},$$

c 由 $P_{1.40}((X_1, \cdots, X_n)\in S)=0.05$ 确定. 查正态分布表，$c=1.645$. 因为

$$\frac{1.41-1.40}{0.04/\sqrt{25}}=1.25<1.645,$$

应当接受 H，即 $a\leqslant 1.40$ 可信.

例 3，例 4 说明对有些零假设或备择假设是复合的检验问题，也有可能找到具有给定水平的一致最强检验. 一般的，对任何检验问题是不是都能找到给定水平的一致最强检验呢？下面的例题回答了这个问题.

【例 6】 $\xi\sim N(a, 1)$，取 $\alpha=0.05$，考虑假设检验问题

$$H:\ a=0;\ K:\ a=-1 \text{ 或 } 1.$$

为简单起见，用容量为 1 的样本 x_1 对(H, K)检验. 由例题 1 知，水平为 0.05 在 $a=1$ 点最强的检验判别区域为 $[1.645, +\infty)$. 可以证明这样的判别区域是唯一的. 所以检验问题 (H, K) 如果有一个一致最强检验，判别区域只能是 $[1.645, +\infty)$. 但它在 $a=-1$ 点不是最强的，因为水平为 0.05 在 $a=-1$ 点最强的唯一检验判别区域是$(-\infty, -1.645]$. 所以检验问题 (H, K) 的水平为 0.05 的一致最强检验是不存在的.

一个检验问题的一致最强检验不存在，是指在一切水平为 α 的检验中不存在一个最好的(确切地说是一致最强的). 对这种情况的处理办法是把要求降低一点. 在水平为 α 的检验集合中不存在最好的，我们就到它的某一个子集中找一个

最好的，这个子集中的检验是符合另外一些附加条件的水平为 α 的检验. 当然，选子集或给检验附加条件应当是"合理"的，要求检验有无偏性就是常用的一个附加条件.

3.3 无偏检验

定义 对假设检验问题 $H: \theta\in\Theta_0$; $K: \theta\in\Theta-\Theta_0$，称判别区域 S 对应的检验为水平为 α 的无偏检验，如果

(1) $P_\theta((x_1, \cdots, x_n)\in S)<\alpha,\ \theta\in\Theta_0$;

(2) $P_\theta((x_1, \cdots, x_n)\in S)\geqslant\alpha,\ \theta\in\Theta-\Theta_0$.

条件(1)要求检验的水平为 α，这不是新的要求. 条件(2)要求不犯第二类错误的概率不小于 α，也就是要求在零假设不真时，拒绝零假设的概率比零假设真时，拒绝零假设的概率"大". 这是一个新的要求，应当说，这个要求是合理的.

在例 6 中取判别区域

$$S=(-\infty, -1.96)\cup(1.96, +\infty)=[x||x|>1.96],$$

由于

$$P_0(X_1\in S)=\int_{-\infty}^{-1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx+\int_{1.96}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$
$$=0.05,$$

$$P_1(X_1\in S)=\int_{-\infty}^{-1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1)^2}dx$$
$$+\int_{1.96}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1)^2}dx$$
$$\geqslant\int_{1.96}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1)^2}dx$$
$$=\int_{0.96}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx\geqslant 0.05,$$

同理，$P_{-1}(X_1\in S)\geqslant 0.05$. 所以，以 S 为判别区域对应的检验是水平为 0.05 的无偏检验.

一般的，若$\xi\sim N(a,\ \sigma^2)$，σ^2已知，则对检验问题

$$H:\ a=a_0;\quad K:\ a\neq a_0$$

来说，判别区域

$$S=\left\{(x_1,\ \cdots,\ x_n)\ \middle|\ |\bar{x}-a_0|>u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}$$

对应的检验（直观的理解是$|\overline{X}-a_0|$太大时拒绝H）是水平为α的无偏检验．其中n是样本的容量，$u_{\frac{\alpha}{2}}$满足

$$\int_{u_{\frac{\alpha}{2}}}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{\alpha}{2},$$

用正态分布表中可以求出．可以证明，这个无偏检验还是一致最强的，即如果还有一个水平为α的无偏检验S'，则

$$P_a((X_1,\ \cdots,\ X_n)\in S)\geqslant P_a((X_1,\ \cdots,\ X_n)\in S'),\ a\neq a.$$

【例7】 已知某炼铁厂在正常情况下铁水含碳量$\xi\sim N(4.55,\ 0.108^2)$．现在测了5炉铁水，其含碳量分别为4.28, 4.40, 4.42, 4.35, 4.37．如果方差没有改变，试问$E\xi$有无变化？

解．需要检验的是$E\xi=4.55$是否可信．我们考虑假设检验问题

$$H:\ E\xi=4.55;\quad K:\ E\xi\neq 4.55.$$

我们已经得到了一个容量为5的样本$x_1,\ x_2,\ x_3,\ x_4,\ x_5$，如信度$\alpha$取0.05，那么判别区域

$$S=\left\{(x_1,\ \cdots,\ x_5)\ \middle|\ \frac{|\bar{x}-4.55|}{1.08/\sqrt{5}}>u_{\frac{0.05}{2}}\right\}$$

对应的检验是水平为0.05的一致最强无偏检验．查正态分布表，$u_{\frac{0.05}{2}}=1.96$．由于

$$\bar{x}=\frac{1}{5}\times(4.28+4.40+4.42+4.35+4.37)=4.364,$$

$$\frac{|\bar{x}-4.55|}{1.08/\sqrt{5}}=\frac{|4.364-4.55|}{1.08/\sqrt{5}}=3.85>1.96,$$

应当拒绝 H，认为 $E\xi$ 有变化.

下面介绍几个常用的，关于正态分布参数检验问题的检验方法，我们将说明它们的最优性质而不给证明.

(1) $\xi\sim N(a,\sigma^2)$，σ^2 未知，考虑检验问题

$$H:\ a\leqslant a_0;\quad K:\ a>a_0, \tag{11}$$

a_0 是指定的数.

在 3.2 例 4 中，我们讨论过 σ^2 已知的情况，那时判别区域取

$$\left\{(x_1,\ \cdots,\ x_n)\left|\frac{\bar{x}-a_0}{\sqrt{\sigma^2/n}}\geqslant c\right.\right\}, \tag{12}$$

即 $\bar{x}-a_0$ 太大时，拒绝 H. 给定信度 α 后，c 由

$$P_{(a_0,\sigma^2)}\left(\frac{\bar{X}-a_0}{\sqrt{\sigma^2/n}}\geqslant c\right)=\alpha$$

决定，从正态分布表可以查到. 对 σ^2 未知的情况仍旧应当在 $\bar{x}-a_0$ 太大时拒绝 H，只是用判别区域(12)时，由于 σ^2 未知，在得到样本点 $(x_1,\ \cdots,\ x_n)$ 后，无法判别它是否属于 S. 现在让(12)中的 σ^2 由它的估计值

$$\frac{n}{n-1}s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

来代替，得

$$\left\{(x_1,\ \cdots,\ x_n)\left|\frac{\bar{x}-a_0}{s}\sqrt{n-1}\geqslant c\right.\right\}. \tag{13}$$

用(13)做判别区域，c 仍由

$$P_{(a_0,\sigma^2)}\left(\frac{\bar{X}-a_0}{s}\sqrt{n-1}\geqslant c\right)=\alpha$$

决定. 注意，在 $a=a_0$ 时，$\frac{\bar{X}-a_0}{s}\sqrt{n-1}$ 不再是标准正态随

机变量了. 由第一节定理 2 知, $\frac{\overline{X}-a_0}{s}\sqrt{n-1}$ 服从自由度为 $n-1$ 的 t 分布. 所以,查 t 分布表可以确定 c.

由于对任何 $a\leqslant a_0$

$$P_{(a,\sigma^2)}\left(\frac{\overline{X}-a_0}{s}\sqrt{n-1}\geqslant c\right)\leqslant P_{(a,\sigma^2)}\left(\frac{\overline{X}-a}{s}\sqrt{n-1}\geqslant c\right)=\alpha$$

以(13)为判别区域对应的检验是(11)的水平为 α 的检验. 可以证明这个检验还是一致最强无偏检验.

(2) $\xi\sim N(a,\ \sigma^2)$, σ^2 未知. 对检验问题

$$H:\ a\geqslant a_0;\quad K:\ a<a_0 \tag{14}$$

来说,应当是 $\bar{x}-a_0$ 太小时拒绝 H. 取判别区域为

$$\left\{(x_1,\ \cdots,\ x_n)\left|\frac{\overline{X}-a_0}{s}\sqrt{n-1}\leqslant c\right.\right\}, \tag{15}$$

c 由
$$P_{(a_0,\sigma^2)}\left(\frac{\overline{X}-a_0}{s}\sqrt{n-1}\leqslant c\right)=\alpha$$

决定, α 是信度. 由于 $a=a_0$ 时, $\frac{\bar{x}-a_0}{s}\sqrt{n-1}$ 服从自由度为 $n-1$ 的 t 分布, c 可以从 t 分布表查到. 可以证明: (15)对应的检验是检验问题(14)的水平为 α 的一致最强无偏检验.

(3) $\xi\sim N(a,\ \sigma^2)$, a 未知. 考虑检验问题

$$H:\ \sigma^2\geqslant\sigma_0^2,\quad K:\ \sigma^2<\sigma_0^2, \tag{16}$$

其中 σ_0^2 是指定的正数.

在样本容量 n 给定后, 在一般情况下, 如果 σ^2 比较大, $ns^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$ 也会比较大. 所以, ns^2 太小时应拒绝 H. 给定信度 α 后,取判别区域为

$$\left\{(x_1,\ \cdots,\ x_n)\left|\frac{ns^2}{\sigma_0^2}<c\right.\right\}, \tag{17}$$

c 由
$$P_{(a,\sigma_0^2)}\left(\frac{ns^2}{\sigma_0^2}<c\right)=\alpha$$

决定. 由于 $\sigma^2=\sigma_0^2$ 时, $\frac{ns^2}{\sigma_0^2}$ 服从自由度为 $n-1$ 的 χ^2 分布(第一节推论 1), c 可以从 χ^2 分布表查到.

由于对任何 $\sigma^2\geqslant\sigma_0^2$,

$$P_{(a,\sigma^2)}\left(\frac{ns^2}{\sigma_0^2}<c\right)\geqslant P_{(a,\sigma^2)}\left(\frac{ns^2}{\sigma^2}<c\right)=\alpha,$$

所以 (17) 对应的检验是一致检验问题 (16) 的水平为 α 的检验. 还可以证明:这个检验是最强无偏检验.

(4) $\xi\sim N(a,\sigma^2)$, a 未知,检验问题

$$H:\ \sigma^2\leqslant\sigma_0^2,\quad K:\ \sigma^2>\sigma_0^2$$

的水平为 α 的一致最强无偏检验的判别区域是

$$\left\{(x_1,\ \cdots,\ x_n)\,\middle|\,\frac{ns^2}{\sigma_0^2}>c\right\}.$$

ns^2 太大时拒绝 H. c 由

$$P_{(a,\sigma_0^2)}\left(\frac{ns^2}{\sigma_0^2}>c\right)=\alpha$$

确定. 可以证明:这个检验还是一致最强的.

3.4 似然比检验法

这是一个构造检验的方法. 设 ξ 有概率分布密度 $p(x;\theta)$(或分布列 $p(a_i;\theta)i=1,\ 2,\ \cdots$), $\theta\in\Theta$, $X_1,\cdots,X_n$ 是 ξ 的一个样本. 考虑检验问题 $H:\ \theta\in\Theta_0\subseteq\Theta$; $K:\ \theta\in\Theta-\Theta_0$. 在 2.1 曾定义过样本的似然函数 $L(x_1,\ \cdots,\ x_n;\ \theta)$. 对每一个样本,视 $L(x_1,\ \cdots,\ x_n;\ \theta)$ 为参数 θ 的函数,若它在 Θ_0 上有极大值, 则记此极大值为 $L(x_1,\ \cdots,\ x_n;\ \hat{\Theta}_0)$, 或 $L(x;\ \hat{\Theta}_0)$; 类似的,记 $L(x_1,\ \cdots,\ x_n;\ \theta)$ 在 Θ 上的极大值为 $L(x_1,\ \cdots,\ x_n;\ \hat{\Theta})$, 或 $L(x;\hat{\Theta})$. 比

$$\frac{L(x_1, \cdots, x_n; \hat{\Theta}_0)}{L(x_1, \cdots, x_n; \hat{\Theta})} \tag{18}$$

是$(x_1, \cdots, x_n)$的函数而不依赖于未知参数θ，记作$\lambda(x_1, \cdots, x_n)$，称其为检验问题$(H, K)$的似然比．因为

$$0 \leqslant L(x_1, \cdots, x_n; \hat{\Theta}_0) \leqslant L(x_1, \cdots, x_n; \hat{\Theta}),$$

$0 \leqslant \lambda(x_1, \cdots, x_n) \leqslant 1$[*]．所谓似然比检验法就是取判别区域为

$$S = \{(x_1, \cdots, x_n) \mid \lambda(x_1, \cdots, x_n) \leqslant c\}$$

的检验方法，常数c由

$$\sup_{\theta \in \Theta_0} P_\theta((X_1, \cdots, X_n) \in S) \leqslant \alpha$$

确定，$0<\alpha<1$是给定的信度．

当$\Theta = \{\theta_0, \theta_1\}$，$\Theta_0 = \{\theta_0\}$时，似然比

$$\lambda(x_1, \cdots, x_n) = \frac{L(x_1, \cdots, x_n; \theta_0)}{\max\limits_{i=0,1} L(x_1, \cdots, x_n; \theta_i)}$$

$$= \begin{cases} 1, \text{当 } L(x_1, \cdots, x_n; \theta_0) \geqslant L(x_1, \cdots, x_n; \theta_1); \\ \dfrac{L(x_1, \cdots, x_n; \theta_0)}{L(x_1, \cdots, x_n; \theta_1)}, \\ \qquad \text{当 } L(x_1, \cdots, x_n; \theta_0) < L(x_1, \cdots, x_n; \theta_1). \end{cases}$$

由于$0<\alpha<1$，只可能有$0<c<1$，

$$S = \left\{(x_1, \cdots, x_n) \,\middle|\, \frac{L(x_1, \cdots, x_n; \theta_0)}{L(x_1, \cdots, x_n; \theta_1)} \leqslant c\right\}$$

$$= \left\{(x_1, \cdots, x_n) \,\middle|\, L(x_1, \cdots, x_n; \theta_1) \geqslant \frac{1}{c} L(x_1, \cdots, x_n; \theta_0)\right\},$$

c由$P_{\theta_0}((x_1, \cdots, x_n) \in S) = \alpha$确定．这和3.2定理1、2的结果一样，在这个意义上可以把似然比检验看作用定理1、2确定检验方法的一种自然推广．似然比检验的直观想法是：当

[*] 使$L(x_1, \cdots, x_n; \hat{\Theta}) = 0$的$(x_1, \cdots, x_n)$出现的概率为0.

零假设真时，(11)的分子和分母接近，所以当似然比太小时拒绝零假设.

下面我们对有关正态分布参数的一些假设检验问题求出似然比检验.

1. t检验：检验均值是否是指定的值

$\xi\sim N(a,\ \sigma^2)$，未知参数$\theta(a,\ \sigma^2)$是二维向量，参数空间$\Theta=\{(a,\ \sigma^2)\,|\,-\infty<a<+\infty,\ \sigma^2>0\}$. 考虑假设检验问题$H$：$a=a_0$；$K$：$a\neq a_0$. 这里$\Theta_0=\{(a,\ \sigma^2)\,|\,a=a_0,\ \sigma^2>0\}$.

为了用似然比方法确定检验，首先要分别求出似然函数

$$L(x;\ \theta)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right\}$$

在Θ_0及Θ上的极大值$L(x;\ \hat{\Theta}_0)$及$L(x;\ \hat{\Theta})$.

用普通求极值的方法不难得到

$$L(x;\ \hat{\Theta}_0)=(2\pi)^{-\frac{n}{2}}\left[\frac{n}{\sum\limits_{i=1}^{n}(x_i-a_0)^2}\right]^{\frac{n}{2}}e^{-\frac{n}{2}}.$$

又在2.1例1中已证明，在Θ上使$L(x;\ \theta)$极大的$a,\ \sigma^2$分别是$\bar{x},\ \frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2=s^2$，所以

$$L(x;\ \hat{\Theta})=(2\pi)^{-\frac{n}{2}}\left(\frac{1}{s^2}\right)^{\frac{n}{2}}e^{-\frac{n}{2}},$$

似然比

$$\begin{aligned}\lambda(x_1,\ \cdots,\ x_n)&=\frac{L(x;\ \hat{\Theta}_0)}{L(x;\ \hat{\Theta})}=\left[\frac{ns^2}{\sum\limits_{i=1}^{n}(x_i-a_0)^2}\right]^{\frac{n}{2}}\\&=\left[\frac{ns^2}{ns^2+n(\bar{x}-a_0)^2}\right]^{\frac{n}{2}}\\&=\left[\frac{1}{1+\frac{(\bar{x}-a_0)^2}{s^2}}\right]^{\frac{n}{2}}.\end{aligned}$$

对于 $0<c<1$, 存在 $t_\alpha \geqslant 0$,

$$S^{*)}=\{\lambda(x_1, \cdots, x_n)\leqslant c\}=\left\{\frac{1}{1+\frac{(\bar{x}-a_0)^2}{s^2}}\leqslant c^{\frac{2}{n}}\right\}$$

$$=\left\{\frac{(\bar{x}-a_0)^2}{s^2}(n-1)\geqslant t_\alpha^2\right\}.$$

为了确定 c 使

$$\sup_{a=a_0} P_{(a,\sigma^2)}(\lambda(X_1, \cdots, X_n)\leqslant c)=\alpha,$$

只需确定 t_α 使

$$\sup_{a=a_0} P_{(a,\sigma^2)}\left(\frac{(\overline{X}-a_0)^2}{s^2}(n-1)\geqslant t_\alpha^2\right)=\alpha$$

或

$$\sup_{a=a_0} P_{(a,\sigma^2)}\left(\frac{|\overline{X}-a_0|}{s}\sqrt{n-1}\geqslant t_\alpha\right)=\alpha,$$

α 是预先给定的信度. 由于 $a=a_0$ 时, 或说 H 真时 $\frac{\overline{X}-a_0}{s}\sqrt{n-1}$ 服从自由度为 $n-1$ 的 t 分布 (见第一节定理 2), 查 t 分布表可以把 t_α 确定下来. 最后得到的水平为 α 的检验是

当 $\frac{|\bar{x}-a_0|}{s}\sqrt{n-1}\geqslant t_\alpha$ 时, 拒绝 H;

当 $\frac{|\bar{x}-a_0|}{s}\sqrt{n-1}<t_\alpha$ 时, 接受 H.

$\int_{t_\alpha}^{+\infty} p_{t_{n-1}}(x)dx=\frac{\alpha}{2}$, $p_{t_{n-1}}(x)$ 是自由度为 $n-1$ 的 t 分布的分布密度.

【例 8】 某糖厂用自动打包机装糖, 每包重量 ξ 服从正态分布, 工作正常时 $E\xi=100$ 斤. 某日开工后测得 9 包重量如下:

---

*) 即 $S=\{(x_1, \cdots, x_n)|\lambda(x_1, \cdots, x_n)\leqslant c\}$.

$$99.3,\ 98.7,\ 100.5,\ 101.2,\ 98.3,$$

$$99.7,\ 99.5,\ 102.1,\ 100.5.$$

试问这一天打包机的工作是否正常?信度取 0.05.

解：$\xi \sim N(a, \sigma^2)$，打包机工作是否正常，即 $a=100$ 是否可信．这个问题可以作为检验问题来处理．考虑检验问题 $H: a=100$; $K: a \neq 100$. 现在已经得到了一个容量为 9 的样本，用 t 检验法：

$$\bar{x}=\frac{1}{9}(99.3+98.7+\cdots+100.5)=99.978,$$

$$s^2=\frac{1}{9}[(99.3-99.978)^2+\cdots+(100.5-99.978)^2]$$

$$=1.306,$$

$$s=1.143,$$

$$\frac{\bar{x}-100}{s}\sqrt{9-1}=-0.054.$$

查 t 分布表，$t_{0.05}=2.31$. 由于

$$\frac{|\bar{x}-100|}{s}\sqrt{8}=0.054<2.31,$$

应当接受 H，即认为打包机工作正常．

2．t 检验：检验两均值是否相等

$\xi \sim N(a, \sigma^2)$，$\eta \sim N(b, \sigma^2)$，a, b, σ^2 是未知参数．ξ 的样本 $X_1, \cdots, X_m$ 与 η 的样本 $Y_1, \cdots, Y_n$ 独立．考虑假设检验问题 $H: a=b$; $K: a \neq b$. 这里

$$\Theta=\{(a, b, \sigma^2) \mid -\infty<a,\ b<+\infty,\ \sigma^2>0\},$$

$$\Theta_0=\{(a, b, \sigma^2) \mid a=b,\ -\infty<a<+\infty,\ \sigma^2>0\}.$$

样本的似然函数

$$L(x,\ y;\ a,\ b,\ \sigma^2)=(2\pi\sigma^2)^{-\frac{m+n}{2}}$$

$$\times\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}(x_i-a)^2+\sum_{j=1}^{n}(y_j-b)^2\right]\right\}$$

$$=(2\pi\sigma^2)^{-\frac{m+n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[ms_x^2+ns_y^2\right.\right.$$

$$\left.\left.+m(\bar{x}-a)^2+n(\bar{y}-b)^2\right]\right\},$$

其中 $\bar{x}=\frac{1}{m}\sum_{i=1}^{m}x_i,\quad \bar{y}=\frac{1}{n}\sum_{j=1}^{n}y_j,$

$$s_x^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\bar{x})^2,\quad s_y^2=\frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})^2.$$

为了使用似然比方法,先求$L(x,\ y;\ \hat{\Theta}_0)$和$L(x,\ y;\ \hat{\Theta})$.

在Θ上,当

$$a=\bar{x},\ b=\bar{y},\ \sigma^2=\hat{\sigma}^2=\frac{1}{m+n}(ms_x^2+ns_y^2)$$

时,$L(x,\ y;a,\ b,\ \sigma^2)$达到极大,这是因为

$$\ln L(x,\ y;\ \bar{x},\ \bar{y},\ \hat{\sigma}^2)-\ln L(x,\ y;\ a,\ b,\ \sigma^2)$$

$$=\left[-\frac{m+n}{2}\ln\frac{ms_x^2+ns_y^2}{m+n}-\frac{m+n}{2}\right]$$

$$-\left[-\frac{m+n}{2}\ln\sigma^2-\frac{ms_x^2+ns_y^2}{2\sigma^2}\right.$$

$$\left.-\frac{m(\bar{x}-a)^2+n(\bar{y}-b)^2}{2\sigma^2}\right]$$

$$\geqslant\frac{m+n}{2}\left[-\ln\frac{ms_x^2+ns_y^2}{(m+n)\sigma^2}-1+\frac{ms_x^2+ns_y^2}{(m+n)\sigma^2}\right]\geqslant 0,$$

所以 $L(x,\ y;\ \hat{\Theta})=(2\pi)^{-\frac{m+n}{2}}\left(\frac{m+n}{ms_x^2+ns_y^2}\right)^{\frac{m+n}{2}}e^{-\frac{m+n}{2}}.$

同样,可以验证,在Θ_0上,当

$$a=b=\hat{a}=\frac{m\bar{x}+n\bar{y}}{m+n},$$

$$\sigma^2=\frac{\sum_{i=1}^{m}(x_i-\hat{a})^2+\sum_{j=1}^{n}(y_j-\hat{a})^2}{m+n}$$

$$=\frac{1}{m+n}\left[ms_x^2+ns_y^2+\frac{mn}{m+n}(\bar{x}-\bar{y})^2\right]$$

时，$L(x,\ y;\ a,\ b,\ \sigma^2)$达到极大，所以

$$L(x,\ y;\ \hat{\theta}_0)=\frac{e^{-\frac{m+n}{2}}}{(2\pi)^{\frac{m+n}{2}}}\left[\frac{m+n}{ms_x^2+ns_y^2+\frac{mn}{m+n}(\bar{x}-\bar{y})^2}\right]^{\frac{m+n}{2}},$$

似然比

$$\lambda(x,\ y)=\left[\frac{ms_x^2+ns_y^2}{ms_x^2+ns_y^2+\frac{mn}{m+n}(\bar{x}-\bar{y})^2}\right]^{\frac{m+n}{2}}$$

$$=\left[\frac{1}{1+\frac{t^2}{m+n-2}}\right]^{\frac{m+n}{2}},$$

其中
$$t^2=\frac{(\bar{x}-\bar{y})^2\Big/\left(\frac{1}{m}+\frac{1}{n}\right)}{(ms_x^2+ns_y^2)/(m+n-2)}.$$

对任何 $0<c<1$，存在 $t_\alpha\geqslant0$，

$$S=\{\lambda(x,\ y)\leqslant c\}=\left\{\frac{1}{1+\frac{t^2}{m+n-2}}\leqslant c^{\frac{2}{m+n}}\right\}$$

$$=\{t^2\geqslant t_\alpha^2\}=\{|t|\geqslant t_\alpha\}.$$

为了确定 c 使 H 真时 $P(\lambda(X,\ Y)<c)=\alpha$，$\alpha$ 是给定的信度，只需确定 t_α，使 H 真时 $P(|t|>t_\alpha)=\alpha$. 注意到

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{ms_x^2+ns_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

在 H 真时服从自由度为 $m+n-2$ 的 t 分布，故查 t 分布表

可以确定 t_α. 最后得到的检验是

$$|t|\geqslant t_\alpha\text{时,拒绝 }H;\quad |t|<t_\alpha\text{时,接受 }H.$$

$\int_{t_\alpha}^{\infty} p_{t_{m+n-2}}(x)dx=\frac{\alpha}{2}$, $p_{t_{m+n-2}}(x)$是自由度为 $m+n-2$ 的 t 分布的分布密度.

【例 9】 某种物品在处理前与处理后取样分析其含脂率如下:

处理前: 0.19, 0.18, 0.21, 0.30, 0.66,
0.42, 0.08, 0.12, 0.30, 0.27;

处理后: 0.15, 0.13, 0.00, 0.07, 0.24, 0.24,
0.19 0.04, 0.08, 0.20, 0.12.

假定处理前后含脂率都服从正态分布,并且方差不变,试问处理前后分布的均值有没有差异,信度取 0.05.

解: 用 ξ, η 分别表示处理前后的含脂率, $\xi\sim N(a,\sigma^2)$, $\eta\sim N(b,\sigma^2)$, 本例的问题是 a, b 是否相等. 作为检验问题来处理,考虑 H: $a=b$; K: $a\neq b$. 现在已经得到容量分别为 $m=10$ 和 $n=11$ 的样本 x 和 y:

$$\bar{x}=(0.19+0.18+\cdots+0.27)\times\frac{1}{10}=0.273,$$

$$\bar{y}=(0.15+0.13+\cdots+0.12)\times\frac{1}{11}=0.133,$$

$$ms_x^2=(0.19-0.273)^2+(0.18-0.273)^2+\cdots+(0.27-0.273)^2=0.253,$$

$$ns_y^2=(0.15-0.133)^2+(0.13-0.133)^2+\cdots+(0.12-0.133)^2=0.064,$$

$$t=\frac{0.273-0.133}{\sqrt{\frac{0.253+0.064}{10+11-2}\left(\frac{1}{10}+\frac{1}{11}\right)}}=2.48.$$

查 t 分布表(自由度为19) $t_{0.05}=2.09$. 由于 $|t|=2.48>2.09$, 应当拒绝 H, 即认为处理后的含脂率降低了.

3. χ^2 检验: 检验方差是否是指定的值

$\xi\sim N(a,\ \sigma^2)$, a, σ^2 未知. 考虑检验问题

$$H:\ \sigma^2=\sigma_0^2;\ K:\ \sigma^2\neq\sigma_0^2,\ \sigma_0^2\text{ 是指定的数.}$$

这里 $\Theta=\{(a,\ \sigma^2)\mid -\infty<a<+\infty,\ \sigma^2>0\}$; $\Theta_0=\{(a,\ \sigma_0^2)\mid -\infty<a<+\infty\}$. ξ 的样本 $X_1,\ \cdots,\ X_n$ 的似然函数

$$L(x;\ a,\ \sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-a)^2\right\},$$

$$L(x;\ \hat{\Theta})=(2\pi)^{-\frac{n}{2}}\left[\frac{n}{\sum_{i=1}^n(x_i-\bar{x})^2}\right]^{\frac{n}{2}}e^{-\frac{n}{2}}.$$

在 Θ_0 上,

$$L(x;\ a,\ \sigma^2)$$
$$=(2\pi\sigma_0^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(x_i-\bar{x})^2-\frac{n}{2\sigma_0^2}(\bar{x}-a)^2\right\},$$

所以在 $a=\bar{x}$ 处, $L(x;\ a,\ \sigma^2)$ 达到极大,

$$L(x;\ \hat{\Theta}_0)=(2\pi\sigma_0^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(x_i-\bar{x})^2\right\}.$$

似然比

$$\lambda(x)=\frac{L(x;\ \hat{\Theta}_0)}{L(x;\ \hat{\Theta})}$$
$$=\left[\frac{\sum_{i=1}^n(x_i-\bar{x})^2}{n\sigma_0^2}\right]^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^n(x_i-\bar{x})^2}{\sigma_0^2}-n\right]\right\}.$$

令 $y(x)=\dfrac{\sum_{i=1}^n(x_i-\bar{x})^2}{\sigma_0^2}$, 则

$$\lambda(x)=\left[\frac{y(x)}{n}\right]^{\frac{n}{2}}e^{-\frac{1}{2}[y(x)-n]}=h[y(x)].$$

求出 $h(y)$ 的导函数可以看到，在$[0, n]$上 $h(y)$ 是增函数，在$[n, +\infty)$上 $h(y)$ 是降函数，在 $y=n$ 处 $h(y)$ 取极大值 1. 因此对任何 $0<c<1$，有常数 c_1、c_2，

$$S=\{\lambda(x)\leqslant c\}=\{y(x)\leqslant c_1 \quad \text{或} \quad y(x)\geqslant c_2\}.$$

给定信度 α 后，为了确定 S 只需确定 c_1、c_2，使得 H 真时

$$P\left(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma_0^2}\leqslant c_1 \quad \text{或} \quad \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma_0^2}\geqslant c_2\right)=\alpha.$$

因为 H 真时，$\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma_0^2}$ 服从自由度为 $n-1$ 的 χ^2 分布，习惯上用查 χ^2 分布表的方法确定 c_1, c_2 使

$$\int_0^{c_1}p_{\chi^2_{n-1}}(x)dx=\int_{c_2}^{+\infty}p_{\chi^2_{n-1}}(x)dx=\frac{\alpha}{2},$$

$p_{\chi^2_{n-1}}(x)$是自由度为 $n-1$ 的 χ^2 分布的分布密度。

最后得到的检验方法是：当

$$c_1\leqslant\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}\leqslant c_2$$

时接受 H，否则拒绝 H.

【例 10】 已知维尼纶纤度在正常条件下服从正态分布 $N(1.405, 0.048^2)$，某日抽取 5 根纤维，测得其纤度为 1.32, 1.55, 1.36, 1.40, 1.44. 问这一天分布方差是否正常？信度取 0.05.

解：用 ξ 表示维尼纶的纤度，$\xi\sim N(a, \sigma^2)$. 本例要检验的统计假设是 H: $\sigma^2=0.048^2$. 用 χ^2 检验法

$$\bar{x}=(1.32+1.55+1.36+1.40+1.44)\times\frac{1}{5}=1.414,$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=(1.32-1.414)^2+(1.55-1.414)^2$$

$$+\cdots+(1.44-1.414)^2=0.0311,$$

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}=\frac{0.0311}{0.048^2}=13.5.$$

查 χ^2 分布表(自由度为4) $c_2<11.67$, 由于 $13.5>11.67$, 应当拒绝假设.

4. F 检验: 检验两个方差是否相等

$\xi\sim N(a, \sigma_1^2)$, $\eta\sim N(b, \sigma_2^2)$, a、b、σ_1^2、σ_2^2 是未知参数. ξ 的样本 $X_1, \cdots, X_m$ 与 η 的样本 $Y_1, \cdots, Y_n$ 独立. 考虑检验问题 H: $\sigma_1^2=\sigma_2^2$; K: $\sigma_1^2\neq\sigma_2^2$. 这里

$$\Theta=\{(a, b, \sigma_1^2, \sigma_2^2)\mid -\infty<a, b<+\infty, \sigma_1^2>0, \sigma_2^2>0\},$$

$$\Theta_0=\{(a, b, \sigma_1^2, \sigma_2^2)\mid -\infty<a, b<+\infty; \sigma_1^2=\sigma_2^2>0\}.$$

样本的似然函数

$$L(x, y; a, b, \sigma_1^2, \sigma_2^2)=(2\pi)^{-\frac{m+n}{2}}(\sigma_1^2)^{-\frac{m}{2}}(\sigma_2^2)^{-\frac{n}{2}}$$

$$\times\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i-a)-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}(y_j-b)^2\right\},$$

在3.3, 2中已证明

$$L(x, y; \hat{\Theta}_0)=(2\pi)^{-\frac{m+n}{2}}\left(\frac{m+n}{ms_x^2+ns_y^2}\right)^{\frac{m+n}{2}}e^{-\frac{m+n}{2}}.$$

用同样的方法可以验证在 Θ 上 $a=\bar{x}$, $b=\bar{y}$, $\sigma_1^2=s_x^2$, $\sigma_2^2=s_y^2$ 时 L 达到极大, 所以

$$L(x, y; \hat{\Theta})=(2\pi)^{-\frac{m+n}{2}}(s_x^2)^{-\frac{m}{2}}(s_y^2)^{-\frac{n}{2}}e^{-\frac{m+n}{2}}.$$

似然比

$$\lambda(x, y)=\frac{L(x, y; \hat{\Theta}_0)}{L(x, y; \hat{\Theta})}=\frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}}n^{\frac{n}{2}}}\frac{(ms_x^2)^{\frac{m}{2}}(ns_y^2)^{\frac{n}{2}}}{(ms_x^2+ns_y^2)^{\frac{m+n}{2}}}$$

$$=\frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}}n^{\frac{n}{2}}}\cdot\frac{\left[\frac{m-1}{n-1}f(x, y)\right]^{\frac{m}{2}}}{\left[1+\frac{m-1}{n-1}f(x, y)\right]^{\frac{m+n}{2}}},$$

其中 $$f(x, y)=\frac{ms_x^2/(m-1)}{ns_y^2/(n-1)}$$

对任何 $0<c<1$，存在 c_1、c_2，

$$S=\{\lambda(x, y)\leqslant c\}=\{f(x, y)\leqslant c_1 \text{ 或 } f(x, y)\geqslant c_2\}.$$

给定信度 $0<\alpha<1$，为确定 c 使 H 真时 $P(\lambda(x, y)\leqslant c)=\alpha$，只需确定 c_1，c_2 使 H 真时

$$P(f(X, Y)\leqslant c_1 \quad \text{或} \quad f(X, Y)\geqslant c_2)=\alpha.$$

由于 H 真时 $f(X, Y)$ 服从第一自由度为 $m-1$，第二自由度为 $n-1$ 的 F 分布，习惯上取 c_1，c_2 使

$$\int_0^{c_1} p_{F_{m-1,n-1}}(x)dx=\int_{c_2}^{+\infty} p_{F_{m-1,n-1}}(x)dx=\frac{\alpha}{2},$$

$p_{F_{m-1,n-1}}(x)$ 是第一自由度为 $m-1$，第二自由度为 $n-1$ 的 F 分布的分布密度．从 F 分布表可以查到 c_1，c_2．最后得到的检验方法是：当

$$c_1<\frac{ms_x^2/(m-1)}{ns_y^2/(n-1)}<c_2$$

时，接受 H，否则拒绝 H．

【例 11】 砖瓦厂有两座砖窑，某日从两窑各取机制红砖若干块，测得抗折强度如下：(单位公斤)

甲窑： 20.51, 25.56, 20.78, 37.27, 36.26, 25.97, 24.62;

乙窑： 32.56, 26.66, 25.64, 33.00, 34.87, 31.03,

如果抗折强度的分布近似于正态分布，试问两窑砖抗折强度的方差有无显著差异，信度取 0.10.

解：分别用 ξ, η 表示甲、乙窑砖的抗折强度，$\xi\sim N(a, \sigma_1^2)$, $\eta\sim N(b, \sigma_2^2)$，要检验的统计假设是 H: $\sigma_1^2=\sigma_2^2$. 用 F 检验法

$$m=7,\quad \frac{ms_x^2}{m-1}=46.7123;$$

$$n=6,\quad \frac{ns_y^2}{n-1}=13.6280;$$

$$f(x, y)=\frac{ms_x^2/(m-1)}{ns_y^2/(n-1)}=\frac{46.7123}{13.6280}=3.43.$$

查 F 分布表(0.05)(第一自由度 N_1 为 6, 第二自由度 N_2 为 5) $c_2=4.95$. 由于 $3.43<c_2$, 应当接受 H, 即认为两窑砖抗折强度的方差没有差异.

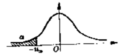
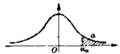
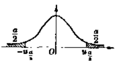
注 由于 F 分布表只能查某些信度 α 对应的 c_2 值，所以算出 $ms_x^2/(m-1)$ 和 $ns_y^2/(n-1)$ 的值后，求 $f(x, y)$ 时要用比较大的值做分子.

关于正态分布参数的假设检验问题还可以提出一些，这里就不一一叙述了. 我们把常用的一些检验方法列成表(表 5-1)，以便于使用.

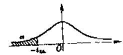
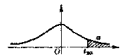
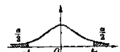
上面我们用比较多的篇幅介绍了有关正态分布参数的假设检验问题及检验方法，这是因为在工农业生产和科学试验中，许多随机现象是用正态随机变量来刻划的，因而人们对正态随机变量的研究也就比较多. 这里最重要的是推导出了许多和正态随机变量有关的统计量的分布. 正是因为知道了这些分布，我们才有可能对给定的，有关正态分布参数的假设检验问题，确定出适当的判别区域.

然而在实际问题中也常常会遇到不是正态随机变量的情

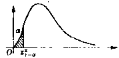
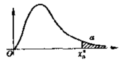
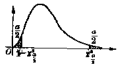
表 5-1

已知条件	检验问题 H　K	统计量	判别区域		统计表
$\xi \sim N(E\xi, \sigma^2)$ σ^2 已知 a 是指定的数 $x_1, x_2, \cdots, x_n$ 是 ξ 的样本	$E\xi \geq a$, $E\xi < a$	$u = \dfrac{\bar{x}-a}{\sigma/\sqrt{n}}$, 其中 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$	$u < -u_\alpha$		正态分布表
	$E\xi \leq a$, $E\xi > a$		$u > u_\alpha$		
	$E\xi = a$, $E\xi \neq a$		$\lvert u\rvert > u_{\frac{\alpha}{2}}$		

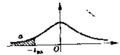
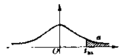
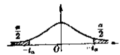
（续表）

已知条件	检验问题 H　K	统计量	判别区域	统计表
$\xi \sim N(E\xi, \sigma^2)$ σ^2 未知 a 是指定的数 $x_1, x_2, \cdots, x_n$ 是 ξ 的样本	$E\xi \geq a$, $E\xi < a$	$t=\dfrac{\bar{x}-a}{s/\sqrt{n-1}}$, 式中 $s=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}$	$t < -t_{2\alpha}$ 	自由度为 $n-1$ 的 t 分布表
	$E\xi \leq a$, $E\xi > a$		$t > t_{2\alpha}$ 	
	$E\xi = a$, $E\xi \neq a$		$\lvert t\rvert > t_\alpha$ 	

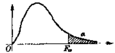
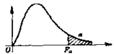
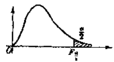
（续表）

已知条件	检验问题 H K	统计量	判别区域		统计表
$\xi\sim N(E\xi, D\xi)$ σ_0^2 是指定的数 $x_1,\cdots,x_n$ 是 ξ 的样本	$D\xi\geqslant\sigma_0^2$, $D\xi<\sigma_0$	$\chi^2=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}$	$\chi^2<\chi_{1-\alpha}^2$		自由度为 $n-1$ 的 χ^2 分布表
	$D\xi\leqslant\sigma_0^2$, $D\xi>\sigma_0$		$\chi^2>\chi_\alpha^2$		
	$D\xi=\sigma_0^2$, $D\xi\neq\sigma_0^2$		$\chi^2<\chi_{1-\frac{\alpha}{2}}^2$ 或 $\chi^2>\chi_{\frac{\alpha}{2}}^2$		

(续表)

已知条件	检验问题 H K	统计量	判别区域		统计表
$\xi \sim N(E\xi, \sigma^2)$ $\eta \sim N(E\eta, \sigma^2)$ σ^2 未知 $x_1, \cdots, x_m$ 是 ξ 的样本 $y_1, \cdots, y_n$ 是 η 的样本	$E\xi \geq E\eta$, $E\xi < E\eta$	$t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{ms_x^2 + ns_y^2}{m+n-2}}\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}$, 其中 $ms_x^2 = \sum_{i=1}^{m}(x_i - \bar{x})^2$, $ns_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$t < -t_{2\alpha}$		自由度为 $m+n-2$ 的 t 分布表
	$E\xi \leq E\eta$, $E\xi > E\eta$		$t > t_{2\alpha}$		
	$E\xi = E\eta$, $E\xi \neq E\eta$		$\|t\| > t_\alpha$		

（续表）

已知条件	检验问题 H　K	统计量	判别区域		统计表
$\xi \sim N(E\xi, D\xi)$ $\eta \sim N(E\eta, D\eta)$ $x_1, \cdots, x_m$ 是 ξ 的样本 $y_1, \cdots, y_n$ 是 η 的样本	$D\xi \geqslant D\eta$, $D\xi < D\eta$	$F = \dfrac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2}{\sum\limits_{i=1}^{m}(x_i - \bar{x})^2} \cdot \dfrac{m-1}{n-1}$	$F > F_\alpha$		第一自由度为 $n-1$ 第二自由度为 $m-1$ 的 F 分布表
	$D\xi \leqslant D\eta$, $D\xi > D\eta$	$F = \dfrac{\sum\limits_{i=1}^{m}(x_i - \bar{x})^2}{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2} \cdot \dfrac{n-1}{m-1}$	$F > F_\alpha$		第一自由度为 $m-1$ 第二自由度为 $n-1$ 的 F 分布表
	$D\xi = D\eta$, $D\xi \neq D\eta$	$F = \dfrac{\max\left\{\dfrac{ms_x^2}{m-1}, \dfrac{ns_y^2}{n-1}\right\}}{\min\left\{\dfrac{ms_x^2}{m-1}, \dfrac{ns_y^2}{n-1}\right\}}$	$F > F_{\frac{\alpha}{2}}$		

况，甚至随机变量的分布是未知的．这时和这个随机变量有关的统计量，如样本平均值的精确分布也是未知的，我们无法根据统计量的精确分布对某一假设检验问题确定适当的判别区域．但是当样本容量无限增大时，有时某些统计量的极限分布是知道的．这时我们有可能利用容量比较大的样本对一些假设检验问题找到适当的判别区域．

3.5 大样本假设检验

1．ξ 是一个随机变量，$D\xi=\sigma_0^2$ 已知，但 ξ 的分布不知道．设想 ξ 取一个容量为可数无穷的样本 $X_1, X_2, \cdots, X_n, \cdots$，这是一个独立同分布的随机变量序列，由中心极限定理 6，对任何实数 x 有

$$\lim_{n\to+\infty} P\left(\frac{\sum_{i=1}^{n} X_i - nE\xi}{\sqrt{n\sigma_0^2}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

或

$$\lim_{n\to+\infty} P\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - E\xi}{\sigma_0/\sqrt{n}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \qquad (19)$$

当 ξ 的分布未知时，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$ 的精确分布也是未知的，但是当 n 比较大时，从(19)知 $\overline{X}$ 的分布接近于 $N(E\xi, \sigma_0^2/n)$．利用这一点有可能对未知参数 $E\xi$ 的某些假设检验问题确定适当的判别区域．例如在 3.2 例 4 中 $\xi \sim N(a, \sigma_0^2)$，对假设检验问题 H: $a \leqslant a_0$，K: $a > a_0$ 确定的判别区域是

$$S=\left\{(x_1, \cdots, x_n) \,\middle|\, \frac{\bar{x}-a_0}{\sigma_0/\sqrt{n}} \geqslant c\right\}, \qquad (20)$$

c 由 $P_{a_0}(S)=\alpha$ 确定．那时使用了一个结论：$a=a_0$ 真时，

$$\frac{\bar{X}-a_0}{\sigma_0/\sqrt{n}}\sim N(0,1), \tag{21}$$

这在 $\xi\sim N(a_0,\ \sigma_0^2)$ 时是对的．ξ 的分布未知时，虽然不能保证(21)成立，但只要样本容量比较大，$\frac{\bar{X}-a_0}{\sigma_0/\sqrt{n}}$ 的分布在 $a=a_0$ 真时接近于 $N(0,\ 1)$ 还是对的．因此仍可以用(20)作判别区域．如已知某种化学纤维强力 ξ 的方差为 1.19^2，希望检验统计假设 H: $E\xi\leqslant 6.2$．取一个容量 n 为100的样本，算得平均值 $\bar{x}=6.35$．在(20)中，当 $\alpha=0.05$ 时，$c=1.645$．因为

$$\frac{6.35-6.2}{1.19/\sqrt{100}}\doteq 1.27<1.645,$$

应当接受 H．在实际问题中如果 $D\xi$ 也未知，因为样本容量比较大，可以用样本的二阶中心矩 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 代替．

2．对多项分布的 χ^2 检验法

设 $\xi_{n1},\ \xi_{n2},\ \cdots,\ \xi_{nr-1}$ 服从参数为 $(n;\ p_1,\ p_2,\ \cdots,\ p_{r-1})$ 的多项分布，$\left(\sum_{i=1}^{r}p_i=1,\ p_i>0,\ i=1,2,\cdots,r\right)$ 即 $(\xi_{n1},\ \cdots,\ \xi_{nr-1})$ 有分布列

$$P(\xi_{n1}=k_1,\ \cdots,\ \xi_{nr-1}=k_{r-1})=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r},$$

$k_1,\ \cdots,\ k_r$ 是非负整数，$\sum_{i=1}^{r}k_i=n$．令 $\xi_{nr}=n-\sum_{i=1}^{r-1}\xi_{ni}$，可以证明*)：对任何实数 x

$$\lim_{n\to+\infty}P\left(\sum_{i=1}^{r}\frac{(\xi_{ni}-np_i)^2}{np_i}<x\right)=\int_{-\infty}^{x}p_{\chi^2_{r-1}}(t)dt, \tag{22}$$

其中 $p_{\chi^2_{r-1}}(t)$ 是自由度为 $r-1$ 的 χ^2 分布的分布密度．

利用这个结果可以对统计假设

*) 证明见复旦大学编《概率论》第二册第一分册(人民教育出版社，1979)．

$$H:\ p_i=p_i^0,\quad i=1,\ 2,\ \cdots,\ r \tag{23}$$

做检验. $p_i^0>0$ 是已知的,满足 $\sum_{i=1}^{r}p_i^0=1$. 基本想法是:如果 H 真,在一般情况下 ξ_{ni} 应当和 np_i^0 比较接近,统计量

$$\sum_{i=1}^{r}\frac{(\xi_{ni}-np_i^0)^2}{np_i^0} \tag{24}$$

应当取比较小的值. 如果(24)取值太大就拒绝 H, 但这又可能犯第一类错误. 我们只能选一个合适的判别区域使犯第一类错误的概率不太大. 给定信度 α, 因为在 n 比较大、H 真时(24)的分布接近于自由度为 $r-1$ 的 χ^2 分布, 存在常数 c 使

$$\int_c^{+\infty}p_{\chi^2_{r-1}}(x)dx=\alpha,$$

就取

$$S=\left\{(k_1,\ \cdots,\ k_r):\sum_{i=1}^{r}\frac{(k_i-np_i^0)^2}{np_i^0}>c\right\} \tag{25}$$

为判别区域.

【例 12】 规定硬币的某一个面为"正面". 为了检验某一个硬币的两个面是否对称, 将硬币投掷了 100 次, 结果有 53 次正面向上, 47 次正面向下. 试问从这个试验结果能否定该硬币的对称性吗?

解: 用 ξ_{n1}, ξ_{n2} 分别表示对硬币投掷 n 次后正面向上、向下的次数, 则 ξ_{n1} 服从参数为 $(n;\ p_1)$ 的二项分布. 令 $p_2=1-p_1$, 硬币两个面对称, 则 $p_1=p_2=\frac{1}{2}$, 所以可以把检验硬币的两个面是否对称的问题看成检验统计假设 $H:\ p_1=\frac{1}{2},\ \ p_2=\frac{1}{2}$ 的问题. 现在 $n=100$, 比较大, 当 H 真时

$$\frac{(\xi_{n1}-np_1)^2}{np_1}+\frac{(\xi_{n2}-np_2)^2}{np_2}$$

的分布接近于自由度为 1 的 χ^2 分布. 取信度 α 为 0.05, 查 χ^2 分布表得 $c=3.84$. 实际观察到的 ξ_{n1} 取 53, ξ_{n2} 取 47, 因为

$$\frac{\left(53-100\times\frac{1}{2}\right)^2}{100\times\frac{1}{2}}+\frac{\left(47-100\times\frac{1}{2}\right)^2}{100\times\frac{1}{2}}=0.36<3.84,$$

应当接受 H, 不能否定该硬币的对称性.

【例 13】 列联表中的独立性检验.

设 A、B 是一个随机试验中的两个随机事件. 如果让你想个办法检验这两个事件是不是独立的, 你怎么做?在第一章中我们曾定义过什么叫事件 A、B 独立, 并且证明了事件 A、B 独立的充分必要条件是 $P(AB)=P(A)P(B)$. 如果你能知道 $P(A)$, $P(B)$, $P(AB)$的值, 那只要看一看 $P(A)P(B)$ 和 $P(AB)$ 是不是相等就行了. 但是在一些实际问题中这三个事件的概率常常是不知道的.

下面我们把这个问题做为一个假设检验问题来解决. 在这个随机试验中, 事件 AB, $A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$ 在每次试验中必出现一个且只出现一个. 用 p_{AB}, $p_{A\bar{B}}$, $p_{\bar{A}B}$, $p_{\bar{A}\bar{B}}$ 及 ξ_{AB}, $\xi_{A\bar{B}}$, $\xi_{\bar{A}B}$, $\xi_{\bar{A}\bar{B}}$ 分别表示事件 AB, $A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$ 的概率及在 n 次独立试验中出现的次数, 则 ξ_{AB}, $\xi_{A\bar{B}}$, $\xi_{\bar{A}B}$ 服从参数为(n; p_{AB}, $p_{A\bar{B}}$, $p_{\bar{A}B}$)的多项分布, 希望检验的统计假设为

$$H:\ p_{AB}=P(A)P(B),\quad p_{A\bar{B}}=P(A)P(\bar{B}),$$
$$p_{\bar{A}B}=P(\bar{A})P(B),\quad p_{\bar{A}\bar{B}}=P(\bar{A})P(\bar{B}). \qquad (26)$$

这个零假设形式上和(23)是一样的, 但是有差别. 在(23)中 p_i^0 是已知的. 现在(26)中的 $P(A)P(B)$, $P(A)P(\bar{B})$,…等等都是未知的. 相应的(24)式变为

$$\frac{[\xi_{AB}-nP(A)P(B)]^2}{nP(A)P(B)}+\frac{[\xi_{A\bar{B}}-nP(A)P(\bar{B})]^2}{nP(A)P(\bar{B})}$$
$$+\frac{[\xi_{\bar{A}B}-nP(\bar{A})P(B)]^2}{nP(\bar{A})P(B)}+\frac{[\xi_{\bar{A}\bar{B}}-nP(\bar{A})P(\bar{B})]^2}{nP(\bar{A})P(\bar{B})},\tag{27}$$

因为 $P(A)$, $P(B)$是未知的, 无法用(27)确定检验 H 的判别区域. 但是可以证明*): 如果 $P(A)$, $P(B)$ 分别用 n 次试验中 A、B 出现的频率代替, 即用$(\xi_{AB}+\xi_{A\bar{B}})/n$ 代替 $P(A)$, 用$(\xi_{AB}+\xi_{\bar{A}B})/n$ 代替 $P(B)$, 当 H 真, n 比较大时,(27)的分布仍接近于 χ^2 分布, 只是自由度不再是 3 而是 1 (从 3 中减去被估计参数的个数 2). 这样在给定信度 α 后, 仍可以确定 c 得到形如(25)的判别区域.

如为了研究患慢性气管炎与吸烟有没有关系, 调查了 339 个 50 岁以上的人,其中吸烟而且患病的有 43 人, 吸烟但未患病的有 162 人. 列成表:

表 5-2

	患慢性气管炎(B)	未患慢性气管炎($\bar{B}$)	合计
吸烟(A)	43	162	205
不吸烟($\bar{A}$)	13	121	134
合计	56	283	

从表上看, 吸烟的 205 人中有 $\frac{43}{205}=21\%$ 的人患病, 而在不吸烟的 134 人中有 $\frac{13}{134}=9.7\%$ 的人患病. 这些统计数字能说明患病的百分比大与吸烟有关系吗? 为了回答这个问题需要检验事件A(被调查的人吸烟)与事件B(被调查的人患

*) 见H·克拉美著,《统计学数学方法》30.5, 魏宗舒等译.

病)是否独立．首先提出零假设 H：事件 A，B 独立．由于 $n=339$ 比较大，可以使用统计量 (27)，取信度 $\alpha=0.01$，查 χ^2 分布表（自由度为 1）得 $c=6.64$．其次计算(27)．用 A 出现的频率 $\frac{205}{339}$ 代替 $P(A)$，用 B 出现的频率 $\frac{56}{339}$ 代替 $P(B)$，得

$$
\begin{aligned}
&\frac{(43-205\times 56/339)^2}{205\times 56/339}+\frac{(162-205\times 283/339)^2}{205\times 283/339}\\
&\quad+\frac{(13-134\times 56/339)^2}{134\times 56/339}+\frac{(121-134\times 283/339)^2}{134\times 283/339}\\
&=\frac{(43-33.86)^2}{33.86}+\frac{(162-171.14)^2}{171.14}\\
&\quad+\frac{(13-22.14)^2}{22.14}+\frac{(121-111.86)^2}{111.86}\\
&=7.47>6.64,
\end{aligned}
$$

应当否定 H，认为 A，B 不独立．

上述检验两个事件是否独立的方法可以推广到检验两组事件是否独立．在一个随机试验中，设事件 A_1, $A_2\cdots$, A_r 在每次试验中必出现一个且仅出现一个，事件 B_1, $B_2\cdots$, B_s 在每次试验中也是必出现一个且仅出现一个．考虑统计假设

$$H:\ P(A_iB_j)=P(A_i)P(B_j),$$

$$i=1,\ 2,\ \cdots,\ r;\quad j=1,\ 2,\ \cdots,\ s.$$

假设在 n 次重复试验中，事件 A_iB_j 出现 n_{ij} 次，$i=1, 2, \cdots, r$；$j=1, 2, \cdots, s$，则

事件 A_i 出现的次数

$$N_{i\cdot}=\sum_{j=1}^{s} n_{ij},\quad i=1,\ 2,\ \cdots,\ r;$$

事件 B_j 出现的次数

$$N_{.j}=\sum_{i=1}^{r} n_{ij}, \quad j=1,\ 2,\ \cdots,\ s.$$

把这些数写成一张表：

表 5-3

	B_1 B_2 $\cdots$ B_s	合　计
A_1	n_{11} n_{12} $\cdots$ n_{1s}	$N_{1.}$
A_2	n_{21} n_{22} $\cdots$ n_{2s}	$N_{2.}$
$\vdots$	$\vdots$ $\vdots$ $\vdots$ $\vdots$	$\vdots$
A_r	n_{r1} n_{r2} $\cdots$ n_{rs}	$N_{r.}$
合　计	$N_{.1}$ $N_{.2}$ $\cdots$ $N_{.s}$	n

称这样一张表为列联表．在

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{[n_{ij}-nP(A_i)P(B_j)]^2}{nP(A_i)P(B_j)}$$

中 $P(A_i)$ 用 A_i 出现的频率 $N_{i.}/n$ 代替 $(i=1,2,\cdots,r)$，$P(B_j)$ 用 B_j 出现的频率 $N_{.j}/n$ 代替得

$$\chi^2=n\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-N_{i.}N_{.j}/n)^2}{N_{i.}N_{.j}}. \tag{28}$$

可以证明在 H 真，n 比较大时，(28)的分布接近于 χ^2 分布，自由度是 $(r-1)(s-1)$（即从 $rs-1$ 中减去被估计的未知参数的个数 $r+s-2$）．给定信度 α 后，查 χ^2 分布表（自由度为 $(r-1)(s-1)$）得到使

$$\int_c^{+\infty} p_{\chi^2_{(r-1)(s-1)}}(x)dx=\alpha$$

的 c．在得到样本 $n_{ij}(1\leqslant i\leqslant r,\ 1\leqslant j\leqslant s)$ 后，计算(28)中 χ^2，若 $\chi^2>c$ 就拒绝 H，否则就接受 H．

【例 14】 下面的数据（表 5-4）是一个自动机床生产的 100 个零件的口径（单位：mm）．

表 5-4

11.02	10.99	10.93	11.01	10.98	10.94	11.02
11.01	10.99	11.00	11.07	10.98	10.97	10.99
10.96	11.02	10.97	10.98	11.00	10.97	11.05
10.95	11.00	11.02	10.99	10.98	11.00	10.98
11.05	11.00	10.98	10.97	11.01	11.07	11.06
10.97	10.94	10.99	11.01	11.03	10.95	11.00
10.95	11.00	11.00	11.00	10.99	10.97	11.02
10.93	10.99	11.02	11.01	11.08	11.00	10.95
11.01	10.93	11.00	10.99	11.01	10.99	10.99
11.04	11.00	11.00	10.97	11.02	11.01	11.04
10.96	10.96	10.99	10.96	10.98	11.08	11.06
10.99	11.00	10.97	11.02	10.98	10.98	11.00
10.99	11.00	11.00	10.99	10.99	10.98	10.97
10.97	10.99	11.03	11.02	11.06	11.04	11.03
11.05	10.96					

按照它们的大小分成 8 组，统计出每组中数据的个数——频数及在全部 100 个数据中所占的比例——频率，列成下表：

表 5-5

组 序	起 止 点	频 数 n_i	频 率 $\frac{n_i}{100}$	累计频率
1	$-\infty$～10.945	5	0.05	0.05
2	10.945～10.965	9	0.09	0.14
3	10.965～10.985	20	0.20	0.34
4	10.985～11.005	33	0.33	0.67
5	11.005～11.025	17	0.17	0.84
6	11.025～11.045	6	0.06	0.90
7	11.045～11.065	6	0.06	0.96
8	11.065～$+\infty$	4	0.04	1.00

还可以用图的形式把各种口径的频率表示出来．取一个直角坐标系，在横坐标轴上标出各组的分点，然后以各组所在的区

间为底，以相应的(频率)/(组距)为高做出一系列矩形如图5-2.通常把这样的图叫频率直方图，或简称直方图.如果用ξ表示自动机床生产一个零件的口径，那么前面列出的100个数据就是ξ的一个容量比较大的样本．现在回过头来看直方图，其中每一个矩形的面积表示ξ在相应区间内取值的频率，一般说，它比较接近于ξ在那个区间内取值的概率，因而直方图的外轮廓线比较接近于ξ的概率分布密度曲线．从直方图看，它有一个峰，比较对称，而且越远离中间值的数出现的频率越小．人们自然会提出这样的问题：(i) ξ是不是服从正态分布？在一些实际问题中甚至可以更明确地提出(ii) ξ是不是服从均值为11，方差是0.03^2的正态分布？

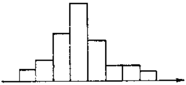


图 5-2

对问题(ii)我们可以算出服从$N(11, 0.03^2)$分布的随机变量在相应的各小区间内取值的概率：

表 5-6

$-\infty$ ~10.945	10.945 ~10.965	10.965 ~10.985	10.985 ~11.005	11.005 ~11.025	11.025 ~11.045	11.045 ~11.065	11.065 ~$+\infty$
0.0336	0.0874	0.1875	0.2590	0.2292	0.1365	0.0518	0.0150
p_1^0	p_2^0	p_3^0	p_4^0	p_5^0	p_6^0	p_7^0	p_8^0

把落在各个小区间内的数据个数n_i($i=1, 2, \cdots, 8$)看作随机变量，则$(n_1, n_2, \cdots, n_7)$服从参数为$(100, p_1, p_2, \cdots, p_7)$的多项分布．检验$\xi$是否服从正态分布$N(11, 0.03^2)$的问题可以作为检验零假设

$$H: \ p_i = p_i^0, \quad i = 1, 2, \cdots, 8$$

的问题来处理.

问题(i)与问题(ii)就有所不同了. 虽然$(n_1, n_2, \cdots, n_7)$仍服从参数为$(100; p_1, p_2, \cdots, p_7)$的多项分布, 检验$\xi$是否服从正态分布的问题仍可以作为检验零假设

$$H: \ p_1 = p_1^0(a, \sigma^2) = \int_{-\infty}^{10.945} p(x; a, \sigma^2)dx,$$

$$p_2 = p_2^0(a, \sigma^2) = \int_{10.945}^{10.965} p(x; a, \sigma^2)dx, \cdots,$$

$$p_8 = p_8^0(a, \sigma^2) = \int_{11.065}^{+\infty} p(x; a, \sigma^2)dx$$

的问题来处理. $p(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}$是$N(a, \sigma^2)$的概率分布密度. 但$p_i^0(i=1, 2, \cdots, 8)$是未知参数的函数, 和例13中碰到的问题一样, 无法用

$$\sum_{i=1}^{8} \frac{[n_i - np_i^0(a, \sigma^2)]^2}{np_i^0(a, \sigma^2)}$$

确定判别区域. 但是可以证明: 如果用样本的平均值$\bar{x}$代替a, 用样本的二阶中心矩s^2代替σ^2, 当H真, n比较大时, 统计量

$$\chi^2 = \sum_{i=1}^{8} \frac{[n_i - np_i^0(\bar{x}, s^2)]^2}{np_i^0(\bar{x}, s^2)} \tag{29}$$

的分布接近于χ^2分布*), 但自由度不是7而是5 (从7中减去被估计的未知参数的个数2). 这样, 我们仍可以把检验ξ是否服从正态分布的问题作为检验统计假设$H: \ p_i = p_i^0(\bar{x}, s^2)$, $i=1, 2, \cdots, 8$的问题来处理. 本例

$$\bar{x} = 10.997, \quad s^2 = 0.033^2,$$

*) 见H·克拉美著《统计学数学方法》30.4, 魏宗舒等译(上海科学技术出版社, 1966).

$$p_1^0(\bar{x}, s^2)=\int_{-\infty}^{10.945}\frac{1}{\sqrt{2\pi}\times 0.033}\times\exp\left\{-\frac{1}{2\times 0.033^2}(x-10.997)^2\right\}dx$$

$$=0.0571,$$

$$p_2^0(\bar{x}, s^2)=0.1089,\quad p_3^0(\bar{x}, s^2)=0.1934,$$

$$p_4^0(\bar{x}, s^2)=0.2354,\quad p_5^0(\bar{x}, s^2)=0.2075,$$

$$p_6^0(\bar{x}, s^2)=0.1242,\quad p_7^0(\bar{x}, s^2)=0.0538,$$

$$p_8^0(\bar{x}, s^2)=0.0197.$$

$$\chi^2=\frac{(5-5.71)^2}{5.71}+\frac{(9-10.89)^2}{10.89}+\frac{(20-19.34)^2}{19.34}+\frac{(33-23.54)^2}{23.54}+\frac{(17-20.75)^2}{20.75}+\frac{(6-12.42)^2}{12.42}+\frac{(6-5.38)^2}{5.38}+\frac{(4-1.97)^2}{1.97}=10.40.$$

查 χ^2 分布表(自由度为 5),若取信度 $\alpha=0.05$, $c=11.07$,由于 $10.40<11.07$, 应当接受 H, 认为 ξ 服从正态分布.

在例 13 及例 14 中所遇到的问题的一般提法是: $\xi_{n1}, \cdots, \xi_{nr-1}$ 服从参数为 $(n; p_1(\alpha_1, \cdots, \alpha_s), \cdots, p_{r-1}(\alpha_1, \cdots, \alpha_s))$ 的多项分布, $\alpha_1, \cdots, \alpha_s$ 是未知参数, $s<r-1$. 令

$$\xi_{nr}=n-\sum_{i=1}^{r-1}\xi_{ni},$$

在某些条件下, $\alpha_i(i=1, 2, \cdots, r)$ 由它的某一估计量 $\hat{\alpha}_i$ 代替后,对任何 $x>0$, 有

$$\lim_{n\to+\infty}P\left(\sum_{i=1}^{r}\frac{[\xi_{ni}-np_i(\hat{\alpha}_1, \cdots, \hat{\alpha}_s)]^2}{np_i(\hat{\alpha}_1, \cdots, \hat{\alpha}_s)}<x\right)$$

$$=\int_{-\infty}^{x}p_{\chi^2_{r-s-1}}(t)dt,$$

其中 $p_{\chi^2_{r-s-1}}(x)$是自由度为 $r-s-1$ 的 χ^2 分布的分布密度*).

3．K 检验法

这一小段要介绍的是检验随机变量 ξ 的分布函数 $F_\xi(x)$ 是否和已知连续分布函数 $F_0(x)$ 相同的另一个检验方法.

设 $x_1, x_2, \cdots, x_n$ 是 ξ 的一个样本，将 $x_1, x_2, \cdots, x_n$ 由小到大排成

$$x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_n^*,$$

定义函数

$$F_n(x)=\begin{cases}0, & \text{当 } x\leqslant x_1^* \text{ 时;}\\ \dfrac{k}{n}, & \text{当 } x_k^*<x\leqslant x_{k+1}^* \text{ 时, } k=1, 2, \cdots, n-1;\\ 1, & \text{当 } x>x_n^* \text{ 时,}\end{cases}$$

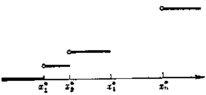


图 5-3

称它为样本的分布函数（或经验分布函数）. $F_n(x)$ 的图象见图 5-3. 对任何指定的 x, $F_n(x)$ 的值就是事件“$\xi<x$”在 n 次试验中出现的频率，是一个离散型随机变量. 柯尔莫哥洛夫证明了**)：当 $F_\xi(x)$是 x 的连续函数时，对任何 z 有

$$\lim_{n\to+\infty} P\left(\sqrt{n}\sup_{-\infty<x<+\infty}|F_n(x)-F_\xi(x)|<z\right)=K(z)$$

$$=\begin{cases}0, & \text{当 } z\leqslant 0 \text{ 时;}\\ \displaystyle\sum_{l=-\infty}^{+\infty}(-1)^l e^{-2l^2z^2}, & \text{当 } z>0 \text{ 时.}\end{cases}$$

$K(z)$的值可以从附表 6 查到.

*) 见 H·克拉美著《统计学数学方法》30.3，魏宗舒等译（上海科学技术出版社，1966）.

**) 这个证明很复杂，想了解证明的读者可以看 M. 费史著《概率论及数理统计》（上海科学技术出版社，1965）的有关章节.

利用上述的柯尔莫哥洛夫极限定理可以检验统计假设 $H: F_\xi(x)=F_0(x)$. 因为 H 真时,对任何指定的 x, $F_0(x)$ 表示事件"$\xi<x$"的概率,注意到 $F_n(x)$ 表示事件"$\xi<x$"在 n 次试验中出现的频率,所以当 n 比较大时,在一般情况下 $F_n(x)$ 和 $F_0(x)$ 的差别是不大的,因而当

$$\sqrt{n}\,D_n=\sqrt{n}\sup_{-\infty<x<+\infty}|F_n(x)-F_0(x)|$$

太大时,应当拒绝 H,否则接受 H. 由于 n 比较大时,$\sqrt{n}D_n$ 的分布接近 $K(z)$,为了控制犯第一类错误的概率,在给定信度 α 后,取 c 使

$$P(\sqrt{n}\,D_n\geqslant c)\approx 1-K(c)=\alpha.$$

在取得样本点后,计算 $\sqrt{n}\,D_n$,若 $\sqrt{n}\,D_n\geqslant c$ 就拒绝 H,否则接受 H.

作为一个例子,我们用 K 检验法检验例 14 中 ξ 是否服从正态分布 $N(11,\ 0.03^2)$.

H: $F_\xi(x)=F_0(x)$

$$=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\times 0.03}\exp\left\{-\frac{1}{2\times 0.03^2}(t-11)^2\right\}dt,$$

显然 $F_0(x)$ 是 x 的连续函数. 对于给定的 n,由于 $F_0(x)$ 是不降函数

$$\begin{aligned}D_n&=\sup_{-\infty<x<+\infty}|F_n(x)-F_0(x)|\\&=\max_{1\leqslant i\leqslant n}\{|F_n(x_i^*)-F_0(x_i^*)|,\ |F_n(x_{i+1}^*)-F_0(x_i^*)|\},\\&\qquad F_n(x_{n+1}^*)-1.\end{aligned}$$

为了计算 D_n,列出表 5-7,可以看到

$$D_n=F_n(11.01)-F_0(11)=0.170,$$

$$\sqrt{n}\,D_n=\sqrt{100}\times 0.170=1.7.$$

取信度 $\alpha=0.05$,查附表 6,$c=1.36$. 由于 $1.7>1.36$,应当

表 5-7

x_i^*	n_i	$F_n(x_i^*)$	$F_0(x_i^*)$	$\|F_n(x_i^*)-F_0(x_i^*)\|$	$F_n(x_{i-1}^*)-F_0(x_i^*)$
10.93	3	0	0.0099	0.0099	0.0201
10.94	2	0.03	0.0228	0.0072	0.0272
10.95	4	0.05	0.0475	0.0025	0.0425
10.96	5	0.09	0.0918	0.0018	0.0482
10.97	10	0.14	0.1587	0.0187	0.0813
10.98	10	0.24	0.2514	0.0114	0.0886
10.99	16	0.34	0.3707	0.0307	0.1293
11.00	17	0.50	0.5000	0.0000	0.1700
11.01	8	0.67	0.6293	0.0407	0.1207
11.02	9	0.75	0.7486	0.0014	0.0914
11.03	3	0.84	0.8413	0.0013	0.0287
11.04	3	0.87	0.9082	0.0382	0.0082
11.05	3	0.90	0.9525	0.0525	0.0225
11.06	3	0.93	0.9772	0.0472	0.0172
11.07	2	0.96	0.9901	0.0301	0.0101
11.08	2	0.98	0.9962	0.0162	0.0038

拒绝 H，不能认为 ξ 服从 $N(11, 0.03^2)$.

习 题 三

1. $\xi \sim N(\theta, 1)$, $x_1, x_2, \cdots, x_9$ 是 ξ 的样本，$\bar{x}=\frac{1}{9}\sum_{i=1}^{9} x_i$.

   (1) 证明下列区域对零假设 H: $\theta=0$ 都是水平为 0.05 的判别区域:

   $s_1=\{(x_1, \cdots, x_9) | 3\bar{x} \geqslant 1.645\}$,

   $s_2=\{(x_1, \cdots, x_9) | 3\bar{x} \leqslant -1.645\}$,

   $s_3=\{(x_1, \cdots, x_9) | 3\bar{x} \leqslant -1.96$ 或 $3\bar{x} \geqslant 1.96\}$,

   $s_4=\{(x_1, \cdots, x_9) | 1.52 \leqslant 3\bar{x} \leqslant 2.19\}$;

   (2) 在上面四个区域中找出对下面检验问题来说是比较好的判别区域.

H_1: $\theta=0$, K_1: $\theta=\theta_1<0$;

H_2: $\theta=0$, K_2: $\theta=\theta_1>0$;

H_3: $\theta=0$, K_3: $\theta\neq0$.

2. ξ 有概率分布密度

$$p(x;\ \theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & 0<x<+\infty;\\ 0, & 其它.\end{cases}$$

(1) 试给出检验问题 H: $\theta=2$, K: $\theta=1$ 的水平为0.05的最强检验方法;

(2) 试给出检验问题 H: $\theta=2$, K: $\theta>2$ 的水平为0.10的一致最强检验方法.

3. 扔骰子300次,结果如下:

点数	1	2	3	4	5	6
频数	43	49	56	45	66	41

根据上列数据能否认为骰子是均匀的. 信度取0.05.

4. 统计一千人的性别与色盲分类如下:

	男	女
正　常	442	514
色　盲	38	6

根据上表检验假设:色盲与性别无关, 信度取0.05.

5. $\xi\sim N(a,\ \sigma^2)$,试用 ξ 的样本: 102, 98, 96, 94, 94, 判断 $a=100$ 是否可信? 信度取5%.

第四节　因子试验的方差分析

4.1 正态随机变量二次型的分布

本节中的假设检验问题将用到正态随机变量二次型的分布, 这要涉及 χ^2 分布. 关于 χ^2 分布已在前面的有关章节中作过一些讨论. 为了便于读者掌握, 现在把主要结论重述

如下.

自由度为 n 的 χ^2 分布指的是由分布密度

$$p(x)=\begin{cases}0, & x\leqslant 0;\\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\end{cases}$$

所确定的概率分布,记作 $\chi^2(n)$(第二章第六节). 若 $\xi\sim\chi^2(n)$(即 ξ 服从 $\chi^2(n)$, 也称 ξ 是自由度为 n 的 χ^2 变量),则

数学期望 $E\xi=n$(第三章第一节例 5, 例 10);

方差 $D\xi=2n$(第三章第二节例 6);

特征函数 $f(t)=(1-2it)^{-\frac{n}{2}}$(第三章第五节例 6).

我们已经得到下面的结论:

1° 若 $Y_1, \cdots, Y_n$ 是独立的标准正态随机变量, 则 $Y_1^2+\cdots+Y_n^2\sim\chi^2(n)$(第二章第六节).

2° 若 $Z_1, \cdots, Z_r$ 是独立的 χ^2 变量, 自由度分别为 $n_1, \cdots, n_r$, 则 $Z_1+\cdots+Z_r\sim\chi^2(n_1+\cdots+n_r)$(第二章第六节习题 11, 第三章第五节例 9).

3° 设 $Y_1, \cdots, Y_n$ 是独立的标准正态随机变量, $Y=(Y_1, \cdots, Y_n)'$, A 是 $n\times n$ 实对称方阵. 若 $Y'AY\sim\chi^2(m)$, 则 A 是幂等阵(即 $A^2=A$)且 A 的秩 $R(A)=m$(第三章第五节例 7).

现在我们进一步证明下面的重要结论.

定理 1(柯赫伦 (Cochran) 定理) $Y_i\sim N(0, 1)$ $(i=1, 2, \cdots, n)$ 独立, $A_1, A_2, \cdots, A_k$ 都是 n 阶对称方阵, 分别有秩 $n_1, n_2, \cdots, n_k$. 记

$$Y=(Y_1, Y_2, \cdots, Y_n)',$$

若

$$Y'Y=Y'A_1Y+Y'A_2Y+\cdots+Y'A_kY,$$

则 $Y'A_iY\sim\chi^2(n_i)(i=1,\ 2,\ \cdots,\ k)$,并且独立的充分必要条件是 $n=n_1+n_2+\cdots+n_k$.

【证】 条件必要性可由前面引述的结论 2° 得出. 这里证明条件的充分性.

A_i 是秩为 n_i 的 n 阶对称方阵, 一定有满秩 n 阶方阵 $(B_i'\ \vdots\ C_i')$, 使

$$A_i=(B_i'\ \vdots\ C_i')\begin{pmatrix}\Lambda_i & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}B_i\\ C_i\end{pmatrix}=B_i'\Lambda_iB_i,$$

其中 B_i 是 n_i 行 n 列矩阵, Λ_i 是 n_i 阶对角阵, 对角线上的元为 1 或 -1, $i=1,\ 2,\ \cdots,\ k$. 令

$$B=(B_1'\ \vdots\ B_2'\ \vdots\ \cdots\ \vdots\ B_k'),$$

$$\Lambda=\begin{pmatrix}\Lambda_1 & & 0\\ & \Lambda_2 & \\ 0 & & \ddots\ \ \Lambda_k\end{pmatrix},$$

因为 $n_1+n_2+\cdots+n_k=n$, B 和 Λ 都是 n 阶方阵,并且

$$\begin{aligned}Y'Y&=Y'B_1'\Lambda_1B_1Y+Y'B_2'\Lambda_2B_2Y+\cdots+Y'B_k'\Lambda_kB_kY\\ &=Y'(B_1'\ \vdots\ B_2'\ \vdots\ \cdots\ \vdots\ B_k')\begin{pmatrix}\Lambda_1 & & 0\\ & \Lambda_2 & \\ 0 & & \ddots\ \ \Lambda_k\end{pmatrix}\begin{pmatrix}B_1\\ B_2\\ \vdots\\ B_k\end{pmatrix}Y\\ &=Y'B'\Lambda BY\end{aligned}$$

对任何实向量 Y 成立,所以 $B'\Lambda B=I_n$(n 阶单位阵), B 是满秩方阵, 有逆矩阵 B^{-1}. 又由于 $\Lambda=(B^{-1})'B^{-1}$ 是正定阵, 对角线上的元不会有 -1, 所以 $\Lambda=I_n$, B 是正交阵. 令

$$Z=(Z_1, Z_2, \cdots, Z_n)'=B(Y_1, Y_2, \cdots, Y_n)'=BY,$$

则 $Z_1, Z_2, \cdots, Z_n$ 仍是独立标准正态随机变量,

$$Y'A_1Y=Y'B_1'A_1B_1Y=Y'B'\begin{pmatrix} I_{n_1} & 0 \\ 0 & 0 \end{pmatrix}BY$$

$$=Z'\begin{pmatrix} I_{n_1} & 0 \\ 0 & 0 \end{pmatrix}Z=\sum_{i=1}^{n_1} Z_i^2,$$

$$Y'A_2Y=Y'B_2'A_2B_2Y=Y'B'\begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & 0 \end{pmatrix}BY$$

$$=\sum_{i=n_1+1}^{n_1+n_2} Z_i^2,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$Y'A_kY=\sum_{i=n_1+\cdots+n_{k-1}+1}^{n_1+\cdots+n_k} Z_i^2,$$

分别是自由度为 $n_1, n_2, \cdots, n_k$ 的 χ^2 变量,并且独立. ▌

引理1 A 是对称幂等方阵(即 $A=A'$ 并且 $A^2=A$), 则 A 的秩 $R(A)$ 等于 A 的迹 $\mathrm{tr}\,A$ *).

【证】 A 对称,一定有正交阵 C, 使

$$A=C\begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & 0 & \\ & & \lambda_r & & \\ & & & 0 & \\ & 0 & & & \ddots \\ & & & & & 0 \end{pmatrix}C',$$

$\lambda_1, \cdots, \lambda_r$ 都不是零, $r=R(A)$. 因为 $A^2=A$, 便有

*) $\mathrm{tr}A=\sum_{i=1}^{n} a_{ii}$, 其中 $A=(a_{ij})$ 为 n 阶方阵; 容易直接验证:

$$\mathrm{tr}(AB)=\mathrm{tr}(BA).$$

$$C\begin{pmatrix}\lambda_1^2 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r^2 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0\end{pmatrix}C'$$

$$=C\begin{pmatrix}\lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0\end{pmatrix}C'C\begin{pmatrix}\lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0\end{pmatrix}C'$$

$$=C\begin{pmatrix}\lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0\end{pmatrix}C'.$$

所以
$$\lambda_1=\cdots=\lambda_r=1,$$

$$A=C\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}C',$$

$$\operatorname{tr}A=\operatorname{tr}ACC'=\operatorname{tr}C'AC=\operatorname{tr}\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}=r=R(A).\quad]$$

引理 2 A, B 是两个对称幂等阵，若 $A-B$ 是非负定阵，则 $A-B$ 幂等，并且 $A-B$ 的秩 $R(A-B)=R(A)-R(B)$.

【证】 A 对称幂等，由引理 1 知，一定有正交阵 C 使

$$A=C\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}C'.$$

设
$$C'BC=\begin{pmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{pmatrix},$$

D_{11} 是 r 阶方阵，则有

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = C'AC = C'BC + C'(A-B)C$$

$$= \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} + \begin{pmatrix} I_r - D_{11} & -D_{12} \\ -D_{21} & -D_{22} \end{pmatrix}.$$

由于 B 对称幂等, $B=BB=B'B$ 是非负定矩阵. 由非负定矩阵的定义知, $C'BC$, $C'(A-B)C$ 都是非负定矩阵, D_{22} 与 $-D_{22}$ 的对角线元非负,因而只能是零, 故 $D_{22}=0$, $D_{12}=0$, $D_{21}=0$. 由

$$\begin{pmatrix} D_{11}^2 & 0 \\ 0 & 0 \end{pmatrix} = (C'BC)(C'BC) = C'BC = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

知, $D_{11}^2 = D_{11}$, 因而 $(I_r - D_{11})^2 = I_r - D_{11}$,

$$C'(A-B)^2C = C'(A-B)CC'(A-B)C$$

$$= \begin{pmatrix} I_r - D_{11} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r - D_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} I_r - D_{11} & 0 \\ 0 & 0 \end{pmatrix} = C'(A-B)C.$$

所以 $(A-B)^2 = A-B$, $A-B$ 幂等. 由引理 1 知

$$R(A-B) = \mathrm{tr}(A-B) = \mathrm{tr}\,A - \mathrm{tr}\,B = R(A) - R(B). \quad ▮$$

用数学归纳法不难证明: 若 A_1, A_2, $\cdots$, A_{k-1} 都是对称幂等阵, A_k 是非负定阵, 并且 $A = \sum_{i=1}^{k} A_i$ 也是幂等阵, 则 A_k 是幂等阵,并且

$$B(A_k) = R(A) - \sum_{i=1}^{k-1} R(A_i).$$

定理 2 $Y_i \sim N(0, 1)$ $(i=1, 2, \cdots, n)$, 独立, A_1, $\cdots$, A_k 都是 n 阶对称方阵. 记 $Y=(Y_1, Y_2, \cdots, Y_n)'$. 若 $Y'A_iY \sim \chi^2(r_i)$, $i=1, 2, \cdots, k-1$, $Y'A_kY$ 是非负二次型,并且

$$Y'AY=\sum_{i=1}^{k}Y'A_iY\sim\chi^2(r),$$

则 $Y'A_kY$ 是 χ^2 变量，自由度为 $r-\sum_{i=1}^{k-1}r_i$，并且 $Y'A_iY\ (i=1, 2, \cdots, k)$，独立.

【证】 $Y'A_iY\sim\chi^2(r_i)$, $i=1, 2, \cdots, k-1$, 利用本节开头引述的结论 3° 知，$A_1, A_2, \cdots, A_{k-1}$ 都是幂等阵，分别有秩 $r_1, r_2, \cdots, r_{k-1}$. $Y'AY\sim\chi^2(r)$, A 幂等，因而 I_n-A 幂等，由引理 1 知

$$R(I_n-A)=\operatorname{tr}(I_n-A)=\operatorname{tr}I_n-\operatorname{tr}A=n-R(A)=n-r,$$

又 $\quad I_n=(I_n-A)+A=(I_n-A)+A_1+\cdots+A_{k-1}+A_k.$

A_k 是非负定阵，由引理 2 知 A_k 的秩为

$$n-(n-r)-\sum_{i=1}^{k-1}r_i=r-\sum_{i=1}^{k-1}r_i.$$

使用定理 1，由于

$$Y'Y=Y'(I-A)Y+Y'A_1Y+\cdots+Y'A_kY,$$

$$R(I-A)+R(A_1)+\cdots+R(A_k)=n,$$

所以 $Y'A_kY\sim\chi^2\left(r-\sum_{i=1}^{k-1}r_i\right)$，并且 $Y'A_iY\ (i=1, 2, \cdots, k)$，独立. 】

4.2 一种方式分组的情形

有 p 个不同的小麦品种 $A_1, A_2, \cdots, A_p$，现在想比较它们在产量方面有没有明显的差别，为此要选一些大小、条件相同的地块，分别种上 $A_1, A_2, \cdots, A_p$，并用同样方法管理. 以后称一个地块为一个小区，设种 A_i 的小区有 n_i 个，产量分别为 $y_{i1}, y_{i2}, \cdots, y_{in_i}\ (i=1, 2, \cdots, p)$，如表 5-8.

我们希望通过对数据 $y_{ij}\ (j=1, 2, \cdots, n_i; i=1, 2, \cdots, p)$

表 5-8

A_1	A_2	$\cdots$	A_p
y_{11}	y_{21}		y_{p1}
y_{12}	y_{22}		y_{p2}
$\vdots$	$\vdots$		$\vdots$
y_{1n_1}	y_{2n_2}		y_{pn_p}

的分析回答 A_1, A_2, $\cdots$, A_p 在产量方面有没有明显的差别.

人们从经验上知道，种品种 A_1 的 n_1 个小区，即使土地、管理等条件完全相同，n_1 个小区的产量 y_{11}, y_{12}, $\cdots$, y_{1n_1} 也不会完全相同，而且有随机性，这是因为从种到收中间还有许多随机因素在起作用.因此应当把 A_1 在一个小区上的产量看作随机变量，记为 ξ_1，而把 y_{11}, y_{12}, $\cdots$, y_{1n_1} 看作 ξ_1 的一个容量为 n_1 的样本 Y_{11}, Y_{12}, $\cdots$, Y_{1n_1} 取的值.类似地，用随机变量 ξ_i 表示 A_i 在一个小区上的产量，把 y_{i1}, y_{i2}, $\cdots$, y_{in_i} 看作 ξ_i 的容量为 n_i 的样本 Y_{i1}, Y_{i2}, $\cdots$, Y_{in_i} 取的值($i=1, 2, \cdots, p$).假设 $\xi_i \sim N(\mu_i, \sigma^2)$, $i=1, 2, \cdots, p$, 并且独立，即 ξ_1, ξ_2, $\cdots$, ξ_p 是有相同方差的独立正态随机变量，数学期望 μ_i 表示在排除了随机干扰这样的理想条件下 A_i 的小区产量，我们就叫它为"理论产量"吧！看 A_1, A_2, $\cdots$, A_p 在产量方面有没有明显差别就看 μ_1, μ_2, $\cdots$, μ_p 是否相同. 我们的问题：A_1, A_2, $\cdots$, A_p 在产量方面有没有明显差别可以作为检验统计假设 H: $\mu_1=\mu_2=\cdots=\mu_p$ 的问题来处理.

把这个问题一般化就是：有 p 个方差相同的独立正态随机变量 ξ_i, $E\xi_i=\mu_i$, $D\xi_i=\sigma^2$, $i=1, 2, \cdots, p$; μ_1, μ_2, $\cdots$, μ_p, σ^2 都是未知的. 将 ξ_i 的样本 $Y_{i1}, \cdots, Y_{in_i}$, $i=1, 2, \cdots, p$, 放在一起构成一个联合样本

$$(Y_{11}, \cdots, Y_{1n_1}; Y_{21}, \cdots, Y_{2n_2}; \cdots; Y_{p1}, \cdots, Y_{pn_p}).$$

现在要用联合样本检验统计假设

$$H:\ \mu_1=\mu_2=\cdots=\mu_p.$$

怎样检验这个统计假设呢？

在一般情况下，$\mu_1, \mu_2, \cdots, \mu_p$ 不全同将反映在

$$Y_{ij}(j=1, \cdots, n_i;\ i=1, 2, \cdots, p)$$

取值的大小不同上，这时离差 $S=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y})^2$ 也比较大，其中 $\overline{Y}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}Y_{ij},\ n=\sum_{i=1}^{p}n_i$. 但是我们还不能只从 S 比较大就断定 $\mu_1, \mu_2, \cdots, \mu_p$ 不全同，因为在 $\mu_1, \mu_2, \cdots, \mu_p$ 全同时，由于抽样的随机性，S 也可能会取较大的值. 为了区分这两种情况，先把离差 S 做一个分解. 令

$$Y_{i\cdot}=\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij},\quad i=1, 2, \cdots, p.$$

$$\begin{aligned}
S&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y})^2\\
&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot}+Y_{i\cdot}-\overline{Y})^2\\
&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2+\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{i\cdot}-\overline{Y})^2\\
&\quad+2\sum_{i=1}^{p}\Big[(Y_{i\cdot}-\overline{Y})\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})\Big]\\
&=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2+\sum_{i=1}^{p}n_i(Y_{i\cdot}-\overline{Y})^2. \qquad (1)
\end{aligned}$$

记上式分解的第一项为 S_e，则

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2.$$

对任一指定的 $1\leqslant i\leqslant p$，$\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2$ 是 ξ_i 的样本 $Y_{i1}, \cdots, Y_{in_i}$ 的离差，是由抽样的随机性造成的. S_e 是一些样本离差

的和,因而也是由抽样的随机性造成的. (1)的第二项

$$\sum_{i=1}^{p} n_i(Y_{i\cdot}-\overline{Y})^2=S_A$$

则不同, 形成 S_A 除了抽样随机性这个原因外, 如果 μ_1, μ_2, $\cdots$, μ_p 不全同,这个差异也要从 S_A 反映出来,一般 S_A 取比较大的值. 因此,和 S_e 比较如果 S_A 不太大,我们只能认为 S_A 只是抽样随机性形成的, 从而接受 H. 和 S_e 比较, 如果 S_A 太大,我们便有理由怀疑 S_A 完全是由抽样随机性形成的,认为 μ_1, μ_2, $\cdots$, μ_p 不全同,从而拒绝 H. 我们将用形如

$$\left\{\frac{S_A}{S_e}>c\right\}$$

的判别区域, c 由预先给定的信度 α 确定. 在给定 α 后,为了确定 c 需要知道统计量 S_A/S_e 在 H 真时的分布.

用 σ^2 除(1)式两端得

$$\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y})^2}{\sigma^2}=\sum_{i=1}^{p}\frac{\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2}{\sigma^2}+\frac{\sum_{i=1}^{p}n_i(Y_{i\cdot}-\overline{Y})^2}{\sigma^2}. \tag{2}$$

右端第一项中的

$$\frac{\sum_{j=1}^{n_i}(Y_{ij}-Y_{i\cdot})^2}{\sigma^2}$$

是自由度为 n_i-1 的 χ^2 变量($i=1, 2, \cdots, p$)(见第一节推论1), 并且这些 χ^2 变量是独立的, 所以第一项 S_e/σ^2 是一些独立的 χ^2 变量之和,仍是 χ^2 变量,自由度是

$$\sum_{i=1}^{p}(n_i-1)=n-p.$$

在 H 真时,(2)式左端是 χ^2 变量,自由度是 $n-1$. 又(2)式右

端第二项非负，根据4.1定理2，在 H 真时

$$\frac{S_A}{\sigma^2}=\frac{\sum_{i=1}^{p}n_i(\bar{Y}_{i\cdot}-\bar{Y})^2}{\sigma^2}$$

是 χ^2 变量，自由度是

$$(n-1)-(n-p)=p-1,$$

而且与 S_e/σ^2 独立，所以在 H 真时

$$\frac{n-p}{p-1}\cdot\frac{S_A}{S_e}=\frac{S_A}{\sigma^2(p-1)}\bigg/\frac{S_e}{\sigma^2(n-p)}$$

是 F 变量，第一自由度是 $p-1$，第二自由度是 $n-p$，判别区域

$$\left\{\frac{S_A}{S_e}>c\right\}=\left\{\frac{n-p}{p-1}\cdot\frac{S_A}{S_e}>\frac{n-p}{p-1}\cdot c\right\}.$$

为了确定 c 或 $(n-p)c/(p-1)$，使

$$P\left(\frac{S_A}{S_e}>c\right)=P\left(\frac{n-p}{p-1}\cdot\frac{S_A}{S_e}>\frac{n-p}{p-1}\cdot c\right)=\alpha,$$

只需从 F 分布表（自由度为 $p-1$, $n-p$）查 $F^\alpha(p-1, n-p)$ 使

$$P(\eta>F^\alpha(p-1, n-p))=\alpha,$$

其中 $\eta\sim F(p-1, n-p)$、最后得到的检验方法是：

若 $\frac{n-p}{p-1}\cdot\frac{S_A}{S_e}>F^\alpha(p-1, n-p)$，就拒绝 H，否则接受 H.

【例1】 电灯泡厂用四种不同材料的灯丝制成四批灯泡，除灯丝材料不同外，其它生产材料完全相同．今由四批灯泡中分别随机抽取数只进行灯泡使用寿命(即一只灯泡能发光的时间，以小时为单位)试验，结果见表5-9．现在希望根据表5-9的数据推断灯泡使用寿命是否因灯丝材料不同而有显著差异，信度用0.05.

表 5-9

灯泡种类			
A_1	A_2	A_3	A_4
1600	1580	1460	1510
1610	1640	1350	1520
1650	1640	1600	1530
1680	1700	1620	1570
1700	1750	1640	1600
1720		1660	1680
1800		1740	
		1820	

解：表 5-9 已给出联合样本 y 的观测值，$n_1=7$，$n_2=5$，$n_3=8$，$n_4=6$，$n=\sum_{i=1}^{4} n_i=26$，$p=4$，查 F 分布表 $F^{0.05}(3,22)=3.05$。下面计算 $\frac{n-p}{p-1}\cdot\frac{S_A}{S_e}$。

令
$$T_i=\sum_{j=1}^{n_i} y_{ij},$$

则
$$y_{i\cdot}=T_i/n_i,\ i=1,\ 2,\ \cdots,\ p,\ \bar{y}=\frac{1}{n}\sum_{i=1}^{p} T_i;$$

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2=\sum_{i=1}^{p}\left(\sum_{j=1}^{n_i} y_{ij}^2-n_i y_{i\cdot}^2\right)$$
$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-\sum_{i=1}^{p}\frac{T_i^2}{n_i};$$

$$S_A=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2-S_e$$
$$=\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-n\bar{y}^2\right)-\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-\sum_{i=1}^{p}\frac{T_i}{n_i}\right)$$
$$=\sum_{i=1}^{p}\frac{T_i^2}{n_i}-\frac{\left(\sum_{i=1}^{p} T_i\right)^2}{n}.$$

利用表 5-10 计算 S_e 与 S_A 是方便的:

表 5-10

灯泡种类	A_1	A_2	A_3	A_4	总和
n_i	7	5	8	6	$n=\sum_{i=1}^{4} n_i=26$
T_i	11760	8310	13090	9410	$\sum_{i=1}^{4} T_i=42570$
T_i^2	138297600	69056100	171348100	88548100	
T_i^2/n_i	19756800	13811220	21418512.5	14758016.7	$\sum_{i=1}^{4} \frac{T_i^2}{n_i}=69744549.2$
$\sum_{j=1}^{n_i} y_{ij}^2$	19783400	13828100	21503700	14778700	$\sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2=69895900$

$$S_e=69895900-69744549.2=151350.8,$$

$$S_A=697445499.2-\frac{42570^2}{26}=44360.7,$$

$$\frac{n-p}{p-1}\frac{S_A}{S_e}=\frac{22}{3}\times\frac{44360.7}{151350.8}=2.15.$$

由于 $2.15<3.05$, 故应当认为灯泡使用寿命不因灯丝材料不同而有差异.

在这个问题中, 四个样本的平均值分别是

$$\bar{y}_{1\cdot}=T_1/7=1680, \qquad \bar{y}_{2\cdot}=T_2/5=1662,$$

$$\bar{y}_{3\cdot}=T_3/7=1636.25, \quad \bar{y}_{4\cdot}=T_4/6=1568.3.$$

如果不做方差分析, 只根据四个平均值不同会做出 A_1 种灯泡使用寿命最长的结论. 但经过方差分析, 发现在使用寿命上四种灯丝制成的灯泡没有显著差别, 这样, 灯泡厂在选择灯丝材料时就可以从其它方面去考虑. 例如在四种灯丝材料中选择能使灯泡成本较低的材料等等.

4.3 两种方式分组的情形

有 p 个不同的小麦品种 $A_1, A_2, \cdots, A_p$，还有 q 个对小麦的不同管理措施(如施肥量、施肥时间) $B_1, B_2, \cdots, B_q$. 现在不仅想比较不同品种在产量上有没有明显的差别，还想比较不同的管理措施对产量有没有明显的影响. 对这个问题当然可以安排两组试验分别检验 $A_1, A_2, \cdots, A_p$ 之间有无明显差异，$B_1, B_2, \cdots, B_q$ 之间有无明显差异. 这里我们介绍一种把两组试验安排在一起的方法，这样做的好处将在后面看到. 选择 pq 块大小、条件相同的地块，仍称地块为小区，为了说起来方便，把这 pq 个小区编上号，分别叫做 $(1, 1)$, $(1, 2)$, $\cdots$, $(1, q)$; $(2, 1)$, $(2, 2)$, $\cdots$, $(2, q)$; $\cdots$; $(p, 1)$, $(p, 2)$, $\cdots$, (p, q). 今在编号为 (i, j) 的小区种上品种 A_i，并采用管理措施 B_j，记得到的产量为 y_{ij} $(i=1, 2, \cdots, p;\ j=1, 2, \cdots, q)$，见表 5-11.

表 5-11

A	B			
	B_1	B_2	$\cdots$	B_q
A_1	y_{11} μ_{11}	y_{12} μ_{12}	$\cdots$	y_{1q} μ_{1q}
A_2	y_{21} μ_{21}	y_{22} μ_{22}	$\cdots$	y_{2q} μ_{2q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_p	y_{p1} μ_{p1}	y_{p2} μ_{p2}		y_{pq} μ_{pq}

以后称品种、管理措施为试验因子，分别用 A、B 表示，称 $A_1, A_2, \cdots, A_p$ 为因子 A 的 p 个水平，$B_1, B_2, \cdots, B_q$ 为因子 B 的 q 个水平.

用 ξ_{ij} 表示对第 i 个品种 A_i 采取第 j 个措施的小区产量. 设 $\xi_{ij}\sim N(\mu_{ij}, \sigma^2)(1\leqslant i\leqslant p;\ 1\leqslant j\leqslant q)$ 独立, μ_{ij} 表示在排除了随机干扰这样的理想条件下, 对品种 A_i 采取管理措施 B_j 后的小区"理论产量", 因子 A、B 对产量的影响就表现在这些 μ_{ij} 的参差不齐上, 或者说表现在 $\mu_{ij}-\mu(1\leqslant i\leqslant p;\ 1\leqslant j\leqslant q)$ 这些数值上,

$$\mu=\sum_{i=1}^{p}\sum_{j=1}^{q}\mu_{ij}/pq$$

是一切 μ_{ij} 的平均值. 有些 μ_{ij} 比平均值 μ 大, 有些 μ_{ij} 比平均值 μ 小, 有的大(小)得很多, 有的大(小)得不多, 差 $\mu_{ij}-\mu$ 反映着 A_i, B_j 对产量的(相对)影响. 对于 $\mu_{ij}-\mu(1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q)$ 还可以做进一步的分析. 对每个 i, 平均 $\mu_{ij}-\mu(1\leqslant j\leqslant q)$, 记 $\mu_{i\cdot}=\dfrac{1}{q}\sum\limits_{j=1}^{q}\mu_{ij}$, 得 $\dfrac{1}{q}\sum\limits_{j=1}^{q}(\mu_{ij}-\mu)=\mu_{i\cdot}-\mu$, 应当说: 如果 $\mu_{1\cdot}-\mu$, $\mu_{2\cdot}-\mu$, $\cdots$, $\mu_{p\cdot}-\mu$ 不全相同, 那只反映 A 的不同水平 A_1, A_2, $\cdots$, A_p 在产量上有差异, 而不反映 B 的不同水平在产量上有差异, 因为 B 的不同水平如果对产量有影响, 这种影响在对 j 平均的过程中也被抵销掉了. 记 $\alpha_i=\mu_{i\cdot}-\mu$, 称它为 A_i 的(主)效应$(1\leqslant i\leqslant p)$. 看 A_1, A_2, $\cdots$, A_p 在产量上有没有差异, 就看 α_1, α_2, $\cdots$, α_p 是否相同, 注意到

$$\sum_{i=1}^{p}\alpha_i=\sum_{i=1}^{p}(\mu_{i\cdot}-\mu)=\frac{1}{q}\sum_{i=1}^{p}\sum_{j=1}^{q}\mu_{ij}-p\mu=0;$$

看 α_1, α_2, $\cdots$, α_p 是否相同, 就看 α_1, α_2, $\cdots$, α_p 是否全为零. 同样记 $\mu_{\cdot j}=\dfrac{1}{p}\sum\limits_{i=1}^{p}\mu_{ij}$, $\beta_j=\mu_{\cdot j}-\mu$, 称 β_j 为因子 B 的第 j、水平 B_j 的(主)效应$(1\leqslant j\leqslant q)$. 看 B_1, B_2, $\cdots$, B_q 在产量上有没有差异, 就看 β_1, β_2, $\cdots$, β_q 是否全为零. 前面说过, $\mu_{ij}-\mu$ 反映着 A_i、B_j 对产量的相对影响, 从其中扣除 A_i、B_j 的效

应，记

$$\gamma_{ij}=(\mu_{ij}-\mu)-\alpha_i-\beta_j, \tag{3}$$

称它为 A_i, B_j 的交互效应（$1\leqslant i\leqslant p$, $1\leqslant j\leqslant q$）. 例如，当品种 A_i 特别适应于管理措施 B_j 时，γ_{ij} 将取比较大的正值；相反的情况，当品种 A_i 特别不适应于管理措施 B_j 时，γ_{ij} 将取绝对值比较大的负值，显然有

$$\sum_{i=1}^{p}\gamma_{ij}=\sum_{i=1}^{p}(\mu_{ij}-\mu)-\sum_{i=1}^{p}\alpha_i-p\beta_j$$
$$=p(\mu_{\cdot j}-\mu)-p\beta_j=0,\quad j=1,\ 2,\ \cdots,\ q;$$
$$\sum_{j=1}^{q}\gamma_{ij}=\sum_{j=1}^{q}(\mu_{ij}-\mu)-q\alpha_i-\sum_{j=1}^{q}\beta_j$$
$$=q(\mu_{i\cdot}-\mu)-q\alpha_i=0,\quad i=1,\ 2,\ \cdots,\ p.$$

从(3)我们得到了 μ_{ij} 的一个分解：

$$\mu_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij},\quad i=1,\ 2,\ \cdots,\ p;\ j=1,\ 2,\ \cdots,\ q.$$

对品种 A_i 采用管理措施 B_j 得到的小区理论产量 μ_{ij} 可以分解为平均值 μ, A_i 的效应 α_i, β_j 的效应 β_j 及 A_i、B_j 的交互效应 γ_{ij} 之和. 视 y_{ij} 为 ξ_{ij} 的容量为 1 的样本 Y_{ij} 取的值，Y_{ij} 有下面的表示式

$$Y_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ij},$$
$$i=1,\ 2,\ \cdots,\ p;\ j=1,\ 2,\ \cdots,\ q. \tag{4}$$

$\varepsilon_{ij}\sim N(0,\ \sigma^2)$, $i=1,\ 2,\ \cdots,\ p$; $j=1,\ 2,\ \cdots,\ q$, 独立. σ^2, μ, α_i, β_j, γ_{ij}（$1\leqslant i\leqslant p$, $1\leqslant j\leqslant q$）是未知参数，满足

$$\sum_{i=1}^{p}\alpha_i=\sum_{j=1}^{q}\beta_j=\sum_{i=1}^{p}\gamma_{ij}=\sum_{j=1}^{q}\gamma_{ij}=0. \tag{5}$$

以上讲的是种品种 A_i 并采用管理措施 B_j 的小区只有一个的情况. 如果把小区增加，使每个品种和每个管理措施搭配的小区都增加到 r 个（$r>1$）就得到 pqr 个小区产量见表 5-12：

表 5-12

A	B			
	B_1	B_2	$\cdots$	B_q
A_1	$y_{111}, y_{112}, \cdots, y_{11r}$	$y_{121}, y_{122}, \cdots, y_{12r}$	$\cdots$	$y_{1q1}, y_{1q2}, \cdots, y_{1qr}$
A_2	$y_{211}, y_{212}, \cdots, y_{21r}$	$y_{221}, y_{222}, \cdots, y_{22r}$	$\cdots$	$y_{2q1}, y_{2q2}, \cdots, y_{2qr}$
A_p	$y_{p11}, y_{p12}, \cdots, y_{p1r}$	$y_{p21}, y_{p22}, \cdots, y_{p2r}$		$y_{pq1}, y_{pq2}, \cdots, y_{pqr}$

y_{ijk} 表示品种 A_i,采用管理措施 B_j,第 k 个重复试验的小区产量($1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q,\ 1 \leqslant k \leqslant r$); 视

$$y_{ij1},\ y_{ij2},\ \cdots,\ y_{ijr}$$

为 ξ_{ij} 的容量为 r 的样本;

$$Y_{ij1},\ Y_{ij2},\ \cdots,\ Y_{ijr}$$

取的值($1 \leqslant i \leqslant p;\ 1 \leqslant j \leqslant q$), Y_{ijk} 有表示式

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$(1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q,\ 1 \leqslant k \leqslant r); \tag{6}$$

$\varepsilon_{ijk} \sim N(0, \sigma^2)$ ($1 \leqslant i \leqslant p;\ 1 \leqslant j \leqslant q;\ 1 \leqslant k \leqslant r$) 独立; σ^2, μ, α_i, β_j, γ_{ij}($1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q$)是满足(5)的未知参数.

我们将从(6)或(4)出发考虑对下面检验问题的检验方法:

(i) H_A: $\alpha_i = 0,\ i = 1,\ 2,\ \cdots,\ p$;

K_A: $\alpha_1, \cdots, \alpha_p$ 中至少有一个不为 0. (7)

(ii) H_B: $\beta_j = 0,\ j = 1,\ 2,\ \cdots,\ q$;

K_B: $\beta_1, \cdots, \beta_q$ 中至少有一个不为 0. (8)

(iii) H_{AB}: $\gamma_{ij} = 0,\ i = 1,\ 2,\ \cdots,\ p;\ j = 1,\ 2,\ \cdots,\ q$;

K_{AB}: $\gamma_{ij},\ i = 1,\ 2,\ \cdots,\ p;\ j = 1,\ 2,\ \cdots,\ q$ 中至少有一个不为 0. (9)

不限于小麦品种和管理措施这样特殊的试验因子和产量这样的特殊指标，按两种方式分组试验问题的一般提法是想考察两个试验因子 A、B 对某一个指标 y 有没有影响，这个影响包括每个试验因子单独对 y 的影响和两个试验因子联合起来对 y 的影响(交互影响)．为此，分别选择 A 和 B 的若干个水平 $A_1, A_2, \cdots, A_p$ 和 $B_1, B_2, \cdots, B_q$，然后在保持其它条件相同的情况下安排试验，观察 y 在 A、B 各种水平搭配下取的值．若每种搭配下的试验重复 r 次($r\geqslant 1$，试验不重复时 $r=1$) 得到 y 取的 pqr 个值见表 5-12，$y_{ijk}(1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q,\ 1\leqslant k\leqslant r)$就可以看作样本(6)($r$ 可以等于 1)取的值．称 α_i 为 A_i 的主效应，β_j 为 B_j 的主效应，γ_{ij} 为 A_i、B_j 的交互效应．我们将从(6)出发检验统计假设(7)、(8)、(9)．

1　交互效应 $\gamma_{ij}(1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q)$为零的情形

这时需要考虑的检验问题只有(7)、(8)，试验不必重复就能对 H_A，H_B 做出拒绝或接受的判断．其检验方法类似于一种方式分组的情形，先对样本 $Y_{ij}(1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q)$的离差做分解：

$$\begin{aligned}
&\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij}-\overline{Y})^2\\
&=\sum_{i=1}^{p}\sum_{j=1}^{q}[(Y_{ij}-Y_{i\cdot}-Y_{\cdot j}+\overline{Y})+(Y_{i\cdot}-\overline{Y})+(Y_{\cdot j}-\overline{Y})]^2\\
&=\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij}-Y_{i\cdot}-Y_{\cdot j}+\overline{Y})^2+q\sum_{i=1}^{p}(Y_{i\cdot}-\overline{Y})^2\\
&\quad+p\sum_{j=1}^{q}(Y_{\cdot j}-\overline{Y})^2=S_e+S_A+S_B,
\end{aligned}$$

其中 $$Y_{i\cdot}=\frac{1}{q}\sum_{j=1}^{q}Y_{ij},\quad Y_{\cdot j}=\frac{1}{p}\sum_{i=1}^{p}Y_{ij},$$

$$\overline{Y}=\frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q}Y_{ij},\quad 1\leqslant i\leqslant p,\ 1\leqslant j\leqslant q,$$

$$S_A=q\sum_{i=1}^{p}(Y_{i\cdot}-\overline{Y})^2,\quad S_B=p\sum_{j=1}^{q}(Y_{\cdot j}-\overline{Y})^2.$$

由于

$$\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij}-\mu-\alpha_i-\beta_j)^2}{\sigma^2}$$

$$=\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij}-Y_{i\cdot}-Y_{\cdot j}+\overline{Y})^2}{\sigma^2}+\frac{q\sum_{i=1}^{p}(Y_{i\cdot}-\overline{Y}-\alpha_i)^2}{\sigma^2}$$

$$+\frac{p\sum_{j=1}^{q}(Y_{\cdot j}-\overline{Y}-\beta_j)^2}{\sigma^2}+\frac{pq(\overline{Y}-\mu)^2}{\sigma^2},$$

并且上式左端 $\sim\chi^2(pq)$，右端第二项 $\sim\chi^2(p-1)$，第三项 $\sim\chi^2(q-1)$，第四项 $\sim\chi^2(1)$，第一项 S_e/σ^2 非负，所以，由4.1节定理2，S_e/σ^2 是 χ^2 变量，自由度是

$$pq-(p-1)-(q-1)-1=(p-1)(q-1),$$

并且右端四个 χ^2 变量独立．所以，

$$\frac{(q-1)q\sum_{i=1}^{p}(Y_{i\cdot}-\overline{Y}-\alpha_i)^2}{S_e}$$

$$=\frac{q\sum_{i=1}^{p}(Y_{i\cdot}-\overline{Y}-\alpha_i)^2/\sigma^2(p-1)}{S_e/\sigma^2(p-1)(q-1)}$$

$$\sim F(p-1,\ (p-1)(q-1))\tag{10}$$

是 F 变量，第一自由度是 $p-1$，第二自由度是 $(p-1)(q-1)$．

$$\frac{(p-1)p\sum_{j=1}^{q}(Y_{\cdot j}-\overline{Y}-\beta_j)^2}{S_e}\sim F(q-1,\ (p-1)(q-1))\tag{11}$$

是 F 变量，第一自由度是 $q-1$，第二自由度是 $(p-1)(q-1)$．

首先考虑检验 H_A 的方法．因为

$$Y_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij},\quad Y_{i\cdot}=\mu+\alpha_i+\varepsilon_{i\cdot},$$
$$\bar{Y}=\mu+\varepsilon_{\cdot\cdot},$$

$Y_{i\cdot}-\bar{Y}=\alpha_i+\varepsilon_{i\cdot}-\varepsilon_{\cdot\cdot}$ 是 α_i 的无偏估计，$\sum\limits_{i=1}^{p}(Y_{i\cdot}-\bar{Y})^2$ 太大时应当否定 H_A，我们采用形如

$$\left\{\frac{(q-1)q\sum\limits_{i=1}^{p}(Y_{i\cdot}-\bar{Y})^2}{S_e}>c\right\} \text{ 或 } \left\{\frac{(q-1)S_A}{S_e}>c\right\}$$

的判别区域．由(10)知，在 H_A 真时，

$$\frac{(q-1)S_A}{S_e}\sim F(p-1,\ (p-1)(q-1)),$$

给定信度 α 后取 $c=F^{\alpha}(p-1,\ (p-1)(q-1))$ 即可．最后得到的检验 H_A 的方法是：

若 $(q-1)S_A/S_e>F^{\alpha}(p-1,\ (p-1)(q-1))$ 就拒绝 H_A，否则接受 H_A．

类似的，检验 H_B 的方法是：

若 $(p-1)S_B/S_e>F^{\alpha}(q-1,\ (p-1)(q-1))$ 就拒绝 H_B，否则接受 H_B．

2　交互作用可能不全为零的情形

对于这种情形，为了检验 H_A，H_B，H_{AB} 时，A、B 各种水平搭配的试验必须有重复．我们从(6)出发，首先分解离差．令

$$Y_{ij\cdot}=\frac{1}{r}\sum_{k=1}^{r}Y_{ijk},\quad Y_{i\cdot\cdot}=\frac{1}{qr}\sum_{j=1}^{q}\sum_{k=1}^{r}Y_{ijk},$$

$$Y_{\cdot j\cdot}=\frac{1}{pr}\sum_{i=1}^{p}\sum_{k=1}^{r}Y_{ijk},\quad \bar{Y}=\frac{1}{pqr}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}Y_{ijk},$$

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(Y_{ijk}-Y_{ij\cdot})^2,\quad S_A=qr\sum_{i=1}^{p}(Y_{i\cdot\cdot}-\bar{Y})^2,$$

$$S_B=pr\sum_{j=1}^{q}(Y_{\cdot j\cdot}-\bar{Y})^2,$$

$$S_{AB}=r\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij.}-Y_{i..}-Y_{.j.}+\bar{Y})^2,$$

则有 $$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(Y_{ijk}-\bar{Y})^2=S_e+S_A+S_B+S_{AB};$$

类似的

$$\frac{pq(r-1)}{p-1}\cdot\frac{qr\sum_{i=1}^{p}(Y_{i..}-\bar{Y}-\alpha_i)^2}{S_e}$$
$$\sim F(p-1,\ pq(r-1));$$

$$\frac{pq(r-1)}{q-1}\cdot\frac{pr\sum_{j=1}^{q}(Y_{.j.}-\bar{Y}-\beta_j)^2}{S_e}$$
$$\sim F(q-1,\ pq(r-1));$$

$$\frac{pq(r-1)}{(p-1)(q-1)}\cdot\frac{r\sum_{i=1}^{p}\sum_{j=1}^{q}(Y_{ij.}-Y_{i..}-Y_{.j.}+\bar{Y}-\tau_{ij})^2}{S_e}$$
$$\sim F((p-1)(q-1),\ pq(r-1)).$$

所以在 H_A 真时

$$F_A=\frac{pq(r-1)}{p-1}\cdot\frac{S_A}{S_e}\sim F(p-1,\ pq(r-1));$$

在 H_B 真时

$$F_B=\frac{pq(r-1)}{q-1}\cdot\frac{S_B}{S_e}\sim F(q-1,\ pq(r-1));$$

在 H_{AB} 真时

$$F_{AB}=\frac{pq(r-1)}{(p-1)(q-1)}\ \frac{S_{AB}}{S_e}$$
$$\sim F((p-1)(q-1),\ pq(r-1)).$$

检验 H_A 的方法是：若 $F_A>F^{\alpha}(p-1,\ pq(r-1))$就拒绝 H_A，否则接受 H_A.

检验 H_B 的方法是：若 $F_B>F^{\alpha}(q-1,\ pq(r-1))$就拒绝 H_B，否则接受 H_B.

检验 H_{AB} 的方法是：若 $F_{AB}>F^{\alpha}((p-1)(q-1),pq(r-1))$就拒绝 H_{AB}，否则接受 H_{AB}.

【例 2】 收缩率(A)和总拉伸倍数(B)是影响合成纤维纤维弹性的两个因素．为了选择好的工艺条件，A、B 各选四个水平做试验，整个试验重复一次，试验结果如下：

表 5-13

B	A: 0 (A_1)	4 (A_2)	8 (A_3)	12 (A_4)
	弹		性	
460 (B_1)	71, 73	73, 75	76, 73	75, 73
520 (B_2)	72, 73	76, 74	79, 77	73, 72
580 (B_3)	75, 73	78, 77	74, 75	70, 71
640 (B_4)	77, 75	74, 74	74, 73	69, 69

本例 $p=4,\ q=4,\ r=2$．取信度 $\alpha=0.01$，查 F 分布表得 $F^{0.01}(3,16)=5.29,\ F^{0.01}(9,16)=3.78$．下面分别计算 $F_A,\ F_B,\ F_{AB}$：令

$$T_{ij\cdot}=\sum_{k=1}^{r}y_{ijk},\quad T_{i\cdot\cdot}=\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk},$$

$$T_{\cdot j\cdot}=\sum_{i=1}^{p}\sum_{k=1}^{r}y_{ijk},\quad T=\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk},$$

则有

$$\begin{aligned}S_A&=qr\sum_{i=1}^{p}(\bar{y}_{i\cdot\cdot}-\bar{y})^2=qr\left(\sum_{i=1}^{p}\bar{y}_{i\cdot\cdot}^2-p\bar{y}^2\right)\\&=qr\sum_{i=1}^{p}\left(\frac{\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk}}{qr}\right)^2-pqr\left(\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk}}{pqr}\right)^2\\&=\sum_{i=1}^{p}\frac{T_{i\cdot\cdot}^2}{qr}-\frac{T^2}{pqr},\end{aligned}$$

$$S_B=\sum_{j=1}^{q}\frac{T_{\cdot j\cdot}^2}{pr}-\frac{T^2}{pqr},$$

$$S_e=\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk}^2-\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij\cdot}^2}{r},$$

$$\begin{aligned}S_{AB}&=\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk}-\bar{y})^2-S_e-S_A-S_B\\&=\frac{1}{r}\sum_{i=1}^{p}\sum_{j=1}^{q}T_{ij\cdot}^2-\frac{1}{qr}\sum_{i=1}^{p}T_{i\cdot\cdot}^2\\&\quad-\frac{1}{pr}\sum_{j=1}^{q}T_{\cdot j\cdot}^2+\frac{1}{pqr}T^2.\end{aligned}$$

在计算 S_e, S_A, S_B, S_{AB} 时，用 $y'_{ijk}=y_{ijk}+c$（c 是任意常数）代替 y_{ijk}（$1\leqslant i\leqslant p$, $1\leqslant j\leqslant q$, $1\leqslant k\leqslant r$），计算结果不变.为了简化计算，从表5-13的试验结果中都减去70得到表5-14.

表 5-14

		A_1	A_2	A_3	A_4	$T_{\cdot j\cdot}$
B_1		1, 3	3, 5	6, 3	5, 3	29
	$T_{i1\cdot}$	4	8	9	8	
B_2		2, 3	6, 4	9, 7	3, 2	36
	$T_{i2\cdot}$	5	10	16	5	
B_3		5, 3	8, 7	4, 5	0, 1	33
	$T_{i3\cdot}$	8	15	9	1	
B_4		7, 5	4, 4	4, 3	−1, −1	25
	$T_{i4\cdot}$	12	8	7	−2	
$T_{i\cdot\cdot}$		29	41	41	12	$T=123$

$$\begin{aligned}S_A&=\frac{1}{8}\times(29^2+41^2+41^2+12^2)-\frac{1}{32}\times123^2\\&=\frac{4347}{8}-\frac{15129}{32}=\frac{2259}{32},\end{aligned}$$

$$S_B=\frac{1}{8}\times(29^2+36^2+33^2+25^2)-\frac{1}{32}\times 123^2$$

$$=\frac{3851}{8}-\frac{15129}{32}=\frac{275}{32},$$

$$S_{AB}=\frac{1}{2}\times[1^2+5^2+\cdots+(-2)^2]$$

$$-\frac{4347}{8}-\frac{3851}{8}+\frac{15129}{32}=\frac{1263}{2}-\frac{17663}{32}$$

$$=\frac{2545}{32},$$

$$S_e=[1^2+3^2+\cdots+(-1)^2]-\frac{1263}{2}$$

$$=653-\frac{1263}{2}=\frac{43}{2},$$

$$F_A=\frac{16}{3}\times\frac{2259}{32}\times\frac{2}{43}=17.51>5.29,$$

$$F_B=\frac{16}{3}\times\frac{275}{32}\times\frac{2}{43}=2.13<5.29,$$

$$F_{AB}=\frac{16}{9}\times\frac{2545}{32}\times\frac{2}{43}=6.57>3.78.$$

方差分析的结果表明 B 的主效应是零，应当选择使 $y_{ij\cdot}$ 最大的 A_i, B_j 作为工艺条件. 从表 5-14 可以看到对应 A_3, B_2 的 $y_{ij\cdot}$ 最大.

习 题 四

1. $\xi=(\xi_1, \xi_2, \cdots, \xi_n)'$ 是 n 个独立的标准正态随机变量, A 是秩为 r 的 n 阶幂等阵, 试证明 $\xi' A\xi\sim\chi^2(r)$.
2. 在一种方式分组模型中, $\mu_i-\mu_j$ 是未知参数 $(1\leqslant i, j\leqslant p)$, 试说明

$$\left(Y_{i\cdot}-Y_{j\cdot}-c\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\frac{S_e}{n-p}},\right.$$

$$\left.Y_{i\cdot}-Y_{j\cdot}+c\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\frac{S_e}{n-p}}\right)\ ..$$

是 $\mu_{i\cdot}-\mu_{j\cdot}$ 的置信系数为 $1-\alpha$ 的区间估计. 其中

$$\int_c^{+\infty} p_{t_{n-p}}(x)dx=\frac{\alpha}{2},$$

$p_{t_{n-p}}(x)$是自由度为 $n-p$ 的 t 分布的分布密度.

3. 对 4 个玉米品种 A_1, A_2, A_3, A_4 用 4 种不同管理方法 B_1, B_2, B_3, B_4 得 16 个小区产量如下表(单位斤). 设对应品种 A_i, 管理方法 B_j 的小区产量 Y_{ij} 可以用(4)式表示, $\gamma_{ij}=0(1\leqslant i, j\leqslant 4)$试检验统计假设

$$H_A:\ \alpha_1=\alpha_2=\alpha_3=\alpha_4=0,\quad H_B:\ \beta_1=\beta_2=\beta_3=\beta_4=0.$$

信度取 5%.

	B_1	B_2	B_3	B_4
A_1	13.5	12.0	14.7	13.2
A_2	15.4	12.9	12.5	12.5
A_3	12.3	12.9	12.0	13.3
A_4	11.5	12.4	11.9	12.3

第五节 线性回归分析

现实世界中的量之间有些有明确的函数关系，如正方形的边长 x 与面积 y 之间就有函数关系 $y=x^2$; 做匀速直线运动的物体的速度 v、运动时间 t 及在时间间隔 t 所走的路程 s 之间有函数关系 $s=vt$. 另外，也有一些量，它们之间没有明确的函数关系，如人的身高与体重，人的年龄与血压等等. 表 5-15 是某平炉 34 炉的熔毕碳(即全部炉料熔化完毕时钢液的含碳量)与精炼时间(从熔毕至出钢冶炼所需要的时间)的生产记录.

把熔毕碳 x 作为横坐标, 精炼时间 y 作为纵坐标, 在坐标纸上画出 34 炉钢熔毕碳与精炼时间对应的点如图 5-4. 从图

表 5-15

编号	熔毕碳 x (0.01%)	精炼时间 y (分)	编号	熔毕碳 x (0.01%)	精炼时间 y (分)
1	180	200	18	116	100
2	104	100	19	123	110
3	134	135	20	151	180
4	141	125	21	110	130
5	204	235	22	108	110
6	150	170	23	158	130
7	121	125	24	107	115
8	151	135	25	180	240
9	147	155	26	127	135
10	145	165	27	115	120
11	141	135	28	191	205
12	144	160	29	190	220
13	190	190	30	153	145
14	190	210	31	155	160
15	161	145	32	177	185
16	165	195	33	177	205
17	154	150	34	143	160

注：这个材料取自中国科学院数学研究所数理统计组编《回归分析方法》。

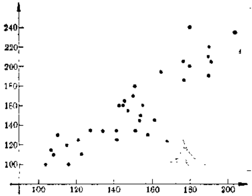


图 5-4

上能看到 y 和 x 之间没有明确的函数关系，但随着 x 的增大 y 有增大的趋势，而且这种趋势是线性的．实际上对于任何一个熔毕碳含量值 x，精炼时间是一个随机变量 Y（实际观测到精炼时间 y 是 Y 所取的值，我们可以用

$$EY=\beta_0+\beta_1 x \tag{1}$$

来刻划 y 和 x 的这种"线性联系"，这里 β_0、β_1 是和 x 无关的常数．由(1)进而把随机变量 Y 表示成

$$Y=\beta_0+\beta_1 x+\varepsilon, \tag{2}$$

$\varepsilon=Y-EY$ 是随机变量，$E\varepsilon=0$．

对 n 炉钢进行独立观测能得到 n 个熔毕碳值 $x_1, x_2, \cdots, x_n$ 及对应的 $y_1, y_2, \cdots, y_n$．因为对任何一个 x_t，在 y_t 取定前不能精确预言它一定取什么值，把 y_t 看作随机变量 Y_t 取的值，而 Y_t 有表示式

$$Y_t=\beta_0+\beta_1 x_t+\varepsilon_t, \quad t=1, 2, \cdots, n, \tag{3}$$

其中 ε_t 表示第 t 次观测中 Y_t 对 $\beta_0+\beta_1 x_t$ 的随机偏差，与 ε 有相同的分布，并且 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 独立．称(3)中相互独立的随机变量 $Y_1, Y_2, \cdots, Y_n$ 为 Y 的样本．和前面一样，有时我们将样本 $Y_1, Y_2, \cdots, Y_n$ 所取的 $y_1, y_2, \cdots, y_n$ 也称为样本．我们希望根据 n 次观测得到的数据 $(x_t; y_t)$，$t=1, 2, \cdots, n$，求出 β_0、β_1 的估计值 b_0、b_1，这样就可以利用方程

$$\hat{y}=b_0+b_1 x, \tag{4}$$

用 x 去估计平均精炼时间 EY 的值．

一般，若随机变量 Y 的数学期望 EY 与变量 $x_1, x_2, \cdots, x_p$ 之间有线性关系

$$EY=\beta_0+\beta_1 x_1+\cdots+\beta_p x_p,$$

则 Y 可以表示成

$$Y=\beta_0+\beta_1 x_1+\cdots+\beta_p x_p+\varepsilon, \tag{5}$$

$\varepsilon=Y-EY$ 是随机变量，$E\varepsilon=0$. 进一步假设 ε 有有限方差 σ^2. σ^2, β_0, β_1, …, β_p 都是与变量 x_1, …, x_p 无关的未知常数，并称 β_i 为 Y 关于 x_i 的**回归系数**，$i=1, 2, \cdots, p$. 对 x_1, x_2, …, x_p 取的 n 组值 $(x_{t1}, x_{t2}, \cdots, x_{tp})$, $t=1, 2, \cdots, n$ 来说，Y 的一个容量为 n 的样本指

$$Y_t=\beta_0+\beta_1x_{t1}+\cdots+\beta_px_{tp}+\varepsilon_t, \quad t=1, 2, \cdots, n, \quad (6)$$

其中 ε_1, ε_2, …, ε_n 是独立随机变量，每个都与 ε 有相同的分布. 我们将从模型(6)出发，根据 n 组值 $(x_{t1}, x_{t2}, \cdots, x_{tp}; y_t)$, $t=1, 2, \cdots, n$, 对未知参数 β_0, β_1, …, β_p, σ^2 进行估计或对与这些参数有关的某些统计假设进行检验.

如果用 $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{Y}$, $\boldsymbol{\varepsilon}$ 分别表示矩阵

$$\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & y_{n2} & \cdots & x_{np} \end{pmatrix}, \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

那么(6)可以表示成矩阵形式

$$\boldsymbol{Y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \quad E\boldsymbol{\varepsilon}=0, \quad D\boldsymbol{\varepsilon}=\sigma^2I_n, \qquad (7)$$

I_n 表示 n 阶单位阵.

在以后的讨论中将比较多地使用矩阵这一工具，否则整个叙述会显得十分繁琐，对矩阵不太熟悉的读者要耐心地、慢慢地读. 另外，今后只讨论 X 是列满秩(即 X 的秩等于它的列数 $p+1$)的情形.

5.1 对 $\boldsymbol{\beta}$ 及 σ^2 的估计

在得到样本 $\boldsymbol{Y}'=(Y_1, Y_2, \cdots, Y_n)$ 后，构造一个 $b'=(b_0, b_1, \cdots, b_p)$ 的函数

$$S(b_0, b_1, \cdots, b_p)=\sum_{t=1}^{n}[Y_t-(b_0+x_{t1}b_1+\cdots+x_{tp}b_p)]^2$$
$$=(\boldsymbol{Y}-Xb)'(\boldsymbol{Y}-Xb), \tag{8}$$

用使 S 达到极小的 $b'=(b_0, b_1, \cdots, b_p)$ 作为 $\beta_0, \beta_1, \cdots, \beta_p$ 的估计值.

从微积分中的极值性质知道，这样的 $b_0, b_1, \cdots, b_p$ 如果存在，必是方程组

$$\frac{\partial S}{\partial b_i}=0, \quad i=0, 1, 2, \cdots, p \tag{9}$$

的解. 把(9)具体写出来并经过整理就得到方程组

$$\begin{aligned}
&nb_0+(\sum_t x_{t1})b_1+(\sum_t x_{t2})b_2+\cdots+(\sum_t x_{tp})b_p=\sum_t Y_t,\\
&(\sum_t x_{t1})b_0+(\sum_t x_{t1}^2)b_1+(\sum_t x_{t1}x_{t2})b_2\\
&\quad+\cdots+(\sum_t x_{t1}x_{tp})b_p=\sum_t x_{t1}Y_t,\\
&(\sum_t x_{t2})b_0+(\sum_t x_{t2}x_{t1})b_1+(\sum_t x_{t2}^2)b_2\\
&\quad+\cdots+(\sum_t x_{t1}x_{tp})b_p=\sum_t x_{t2}Y_t,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&(\sum_t x_{tp})b_0+(\sum_t x_{tp}x_{t1})b_1+(\sum_t x_{tp}x_{t2})b_2\\
&\quad+\cdots+(\sum_t x_{tp}^2)b_p=\sum_t x_{tp}Y_t.
\end{aligned} \tag{10}$$

用矩阵表示就是

$$X'Xb=X'\boldsymbol{Y}, \tag{11}$$

$X'X$ 是满秩方阵(秩 $X'X$ = 秩 $X=p+1$)，所以方程组(10)或(11)有唯一解

$$\tilde{b}=(X'X)^{-1}X'\boldsymbol{Y}. \tag{12}$$

解(12)能不能使(8)达到极小呢?对任何 $b'=(b_0, b_1, \cdots, b_p)$,

$$
\begin{aligned}
S(b_0, b_1, \cdots, b_p) &= (\boldsymbol{Y}-Xb)'(\boldsymbol{Y}-Xb) \\
&= [(\boldsymbol{Y}-X\tilde{b})+(X\tilde{b}-Xb)]'[(\boldsymbol{Y}-X\tilde{b})+(X\tilde{b}-Xb)] \\
&= (\boldsymbol{Y}-X\tilde{b})'(\boldsymbol{Y}-X\tilde{b})+(\boldsymbol{Y}-X\tilde{b})'(X\tilde{b}-Xb) \\
&\quad +(X\tilde{b}-Xb)'(\boldsymbol{Y}-X\tilde{b})+(X\tilde{b}-Xb)'(X\tilde{b}-Xb).
\end{aligned}
$$

等式右边第二项

$$
\begin{aligned}
(\boldsymbol{Y}-X\tilde{b})'X(\tilde{b}-b) &= [X'(\boldsymbol{Y}-X\tilde{b})]'(\tilde{b}-b) \\
&= (X'\boldsymbol{Y}-X'X\tilde{b})'(\tilde{b}-b)=0,
\end{aligned}
$$

同样第三项也是零，第四项是列向量 $X\tilde{b}-Xb$ 长度的平方，非负，故有

$$S(b_0, b_1, \cdots, b_p) \geqslant (\boldsymbol{Y}-X\tilde{b})'(\boldsymbol{Y}-X\tilde{b}),$$

所以(12)能使(8)达到极小.

称方程组(10)或(11)为回归模型(6)或(7)的正规方程，称(12)为 β 的最小二乘估计，记作 $\hat{\beta}=(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)'$，称(8)的极小值 $S(\hat{\beta})$ 为剩余平方和. 若记

$$\hat{Y}_t=\hat{\beta}_0+\hat{\beta}_1x_{t1}+\cdots+\hat{\beta}_px_{tp}, \quad t=1, 2, \cdots, n,$$

则

$$S(\hat{\beta})=\sum_t(Y_t-\hat{Y}_t)^2=(\boldsymbol{Y}-X\hat{\beta})'(\boldsymbol{Y}-X\hat{\beta}); \tag{13}$$

又称函数

$$y=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_px_p \tag{14}$$

为 Y 关于变量 $x_1, x_2, \cdots, x_p$ 的回归方程.

在实际应用中也常使用另一个方法计算 $\hat{\beta}$.

设 $\beta_{(1)}=(\beta_1, \beta_2, \cdots, \beta_p)'$，$\hat{\beta}_{(1)}=(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p)'$，它们分别是 β 和 $\hat{\beta}$ 的后 p 个分量组成的列向量，则

$$\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \beta_{(1)}\end{pmatrix}, \quad \hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_{(1)}\end{pmatrix}.$$

前面已经证明，$\hat{\beta}=(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)'$ 是使

$$S(b)=\sum_t(Y_t-b_0-b_1x_{t1}-\cdots-b_px_{tp})^2$$

取极小值的唯一向量，但是从另一方面看

$$\begin{aligned}S(b)&=\sum_t[(Y_t-\bar{Y})-b_1(x_{t1}-\bar{x}_1)-\cdots-b_p(x_{tp}-\bar{x}_p)\\&\quad+(\bar{Y}-b_0-b_1\bar{x}_1-\cdots-b_p\bar{x}_p)]^2\\&=\sum_t[(Y_t-\bar{Y})-b_1(x_{t1}-\bar{x}_1)-\cdots-b_p(x_{tp}-\bar{x}_p)]^2\\&\quad+n(\bar{Y}-b_0-b_1\bar{x}_1-\cdots-b_p\bar{x}_p)^2\\&=\tilde{S}(b_1,\ b_2,\ \cdots,\ b_p)+n(\bar{Y}-b_0-b_1\bar{x}_1-\cdots-b_p\bar{x}_p)^2,\end{aligned}\tag{15}$$

其中 $\bar{Y}=\frac{1}{n}\sum_t Y_t,\ \bar{x}_i=\frac{1}{n}\sum_t x_{ti},\ i=1,\ 2,\ \cdots,\ p.$

(15)式右端第二项非负，如有 $b_1',\ b_2',\ \cdots,\ b_p'$ 使 $\tilde{S}$ 极小，则再取

$$b_0'=\bar{Y}-b_1'\bar{x}_1-\cdots-b_p'\bar{x}_p,\tag{16}$$

$(b_0',\ b_1',\ \cdots,\ b_p')$就能使 $S(b)$ 极小，因而

$$\hat{\beta}=(\hat{\beta}_0,\ \hat{\beta}_1,\ \cdots,\ \hat{\beta}_p)=(b_0',\ b_1',\ \cdots,\ b_p').$$

令

$$\tilde{X}=\begin{bmatrix}x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p\\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p\end{bmatrix},$$

$$\tilde{Y}=\begin{pmatrix}Y_1-\bar{Y}\\ Y_2-\bar{Y}\\ \vdots\\ Y_n-\bar{Y}\end{pmatrix},\quad 1_n=\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix},$$

由于 X 列满秩，

$$(1_n\ \vdots\ \tilde{X})=X\begin{pmatrix}1 & -\bar{x}_1 & \cdots & -\bar{x}_p\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}$$

仍列满秩，因而 $\tilde{X}$ 列满秩，方程组

$$\widetilde{X}'\widetilde{X}\begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix}=\widetilde{X}'\widetilde{Y} \tag{17}$$

有唯一解 $(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{Y}$，这个解使 $\tilde{S}(b_1, \cdots, b_p)$ 达到极小. 再令

$$\begin{aligned} l_{ij}&=\sum_t(x_{ti}-\bar{x}_i)(x_{tj}-\bar{x}_j), \quad 1\leqslant i,\ j\leqslant p; \\ l_{iy}&=\sum_t(x_{ti}-\bar{x}_i)(Y_t-\overline{Y}), \quad 1\leqslant i\leqslant p. \end{aligned} \tag{18}$$

方程组(17)可表示成

$$\begin{aligned} &l_{11}b_1+l_{12}b_2+\cdots+l_{1p}b_p=l_{1y}, \\ &l_{21}b_1+l_{22}b_2+\cdots+l_{2p}b_p=l_{2y}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &l_{p1}b_1+l_{p2}b_2+\cdots+l_{pp}b_p=l_{py}. \end{aligned} \tag{19}$$

解方程组(19)，然后再用(16)确定 $\hat{\beta}_0$，这是常用的计算 $\hat{\beta}$ 的方法. 这个方法的好处是和(10)比较，方程组(19)包含的未知数少了一个，解方程组的计算量少一些.

$p=1$ 的情况在实际中是经常遇到的，它的结果可以更简洁地表述如下：

设 $\quad X'=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}, \quad \bar{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i,$

则有 $\quad \widetilde{X}'=(x_1-\bar{x}, \cdots, x_n-\bar{x}),$

$$l_{11}=\sum_t(x_t-\bar{x})^2, \quad l_{1y}=\sum_t(x_t-\bar{x})(Y_t-\overline{Y}),$$

正规方程为 $\quad l_{11}b_1=l_{1y},$

$$\hat{\beta}_1=\frac{l_{1y}}{l_{11}}=\frac{\sum_t(x_t-\bar{x})(Y_t-\overline{Y})}{\sum_t(x_t-\bar{x})^2}, \quad \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\bar{x}. \tag{20}$$

【例 1】 试用表5-15中34炉平炉熔毕碳含量与精炼时间的生产记录，求精炼时间 Y 对熔毕碳含量 x 的回归方程.

解：

$$X'=\begin{pmatrix}1 & 1 & \cdots & 1\\ 180 & 104 & \cdots & 143\end{pmatrix},\quad y'=(200\ \ 100\ \ \cdots\ \ 160),$$

正规方程为
$$X'X\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}=X'y,$$

即
$$34\hat{\beta}_0+5103\hat{\beta}_1=5380,$$
$$5103\hat{\beta}_0+791363\hat{\beta}_1=839800.$$

解得
$$\hat{\beta}_0=-32.38,\quad \hat{\beta}_1=1.27.$$

另法

$$\bar{x}=\frac{1}{34}\times(180+104+\cdots+143)=150.09,$$

$$\bar{y}=\frac{1}{34}\times(200+100+\cdots+160)=158.23,$$

$$l_{11}=25462.7,\quad l_{1y}=32325.3,$$

$$\hat{\beta}_1=\frac{l_{1y}}{l_{11}}=1.27,\quad \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=-32.38,$$

Y 关于 x 的回归方程是

$$y=\hat{\beta}_0+\hat{\beta}_1x=-32.38+1.27x. \tag{21}$$

当我们知道了回归方程(21)后，就可以用熔毕碳含量 x 来预测平均精炼时间 EY. 例如这个平炉某一次熔毕碳含量 x 为 150，那么可以预测平均精炼时间

$$EY=-32.28+1.27\times 150=158.12.$$

如果 $Y\sim N(\alpha+\beta x,\ \sigma^2)$，而且 σ 已知，那么在知道 x 后还可以对 EY 做区间估计，例如大体上有

$$P(\hat{\beta}_0+\hat{\beta}_1x-1.96\sigma<EY<\hat{\beta}_0+\hat{\beta}_1x+1.96\sigma)=0.95,$$

$$P(\hat{\beta}_0+\hat{\beta}_1x-3\sigma<EY<\hat{\beta}_0+\hat{\beta}_1x+3\sigma)=0.997,$$

等等. 但 σ^2 一般也是未知的，对它也需要估计，在线性回归分析的一般模型(7)中，当 σ^2 未知时，我们将用

$$\frac{S(\hat{\beta})}{n-p-1}=\frac{1}{n-p-1}(\boldsymbol{Y}-X\hat{\beta})'(\boldsymbol{Y}-X\hat{\beta})$$
$$=\frac{1}{n-p-1}[\boldsymbol{Y}'\boldsymbol{Y}-(X'\boldsymbol{Y})'\hat{\beta}], \qquad (22)$$

或

$$\frac{S(\hat{\beta})}{n-p-1}=\frac{S(\hat{\beta}_{(1)})}{n-p-1}$$
$$=\frac{1}{n-p-1}(\tilde{\boldsymbol{Y}}-\tilde{X}\hat{\beta}_{(1)})'(\tilde{\boldsymbol{Y}}-\tilde{X}\hat{\beta}_{(1)})$$
$$=\frac{1}{n-p-1}[\tilde{\boldsymbol{Y}}'\tilde{\boldsymbol{Y}}-(\tilde{X}'\tilde{\boldsymbol{Y}})'\hat{\beta}_{(1)}]$$
$$=\frac{1}{n-p-1}\left[\sum_t(Y_t-\overline{Y})^2-\sum_{i=1}^{p}l_{iy}\hat{\beta}_i\right] \qquad (23)$$

作为σ^2的估计量，记作$\hat{\sigma}^2$.

特别，当$p=1$时，

$$\frac{S(\hat{\beta})}{n-p-1}=\frac{1}{n-2}[\sum_t(Y_t-\overline{Y})^2-l_{1y}\cdot\hat{\beta}_1]$$
$$=\frac{1}{n-2}[\sum_t(Y_t-\overline{Y})^2-\hat{\beta}_1^2\sum_t(x_t-\bar{x})^2]. \qquad (24)$$

【例 2】 仍用表 5 15 中 34 炉平炉熔毕碳含量与精炼时间的生产记录估计精炼时间 Y 的方差 σ^2.

解：在例 1 中已算出 $\hat{\beta}_1=1.27$, $l_{1y}=32325.3$, 由(24)知

$$\hat{\sigma}^2=\frac{1}{34-2}\times[\sum_t(y_t-\bar{y})^2-32325.3\times1.27]$$
$$=\frac{1}{32}\times[50094.12-41053.13]$$
$$=\frac{1}{32}\times9041=282.53.$$

5.2 $\hat{\beta}$, $\hat{\sigma}^2$ 的性质

性质 1 $\hat{\beta}$ 是 β 的无偏估计.

【证】 $E\hat{\beta}=E(X'X)^{-1}X'\boldsymbol{Y}$

$$=(X'X)^{-1}X'E\boldsymbol{Y}$$

$$=(X'X)^{-1}X'X\beta=\beta. \tag{25}$$

由性质1知 $E\hat{\beta}_i=\beta_i$, $i=1, 2, \cdots, p$, 每个 $\hat{\beta}_i$ 都是 β_i 的无偏估计. ▌

性质 2 设 l 是 $p+1$ 维列向量, 则 $l'\hat{\beta}$ 是 $l'\beta$ 的无偏估计, 并且在 $l'\beta$ 的一切形如 $a'\boldsymbol{Y}=a_1Y_1+\cdots+a_nY_n$ 的无偏估计(即线性无偏估计)中, $l'\hat{\beta}$ 有最小方差.

【证】 $El'\hat{\beta}=l'E\hat{\beta}=l'\beta$, 所以 $l'\hat{\beta}$ 是 $l'\beta$ 的无偏估计. 若 $a'\boldsymbol{Y}$ 也是 $l'\beta$ 的无偏估计, 根据无偏估计的定义

$$a'X\beta=a'E\boldsymbol{Y}=Ea'\boldsymbol{Y}=l'\beta$$

对一切 β 成立, 所以有 $a'X=l'$.

$$D(a'\boldsymbol{Y})=D(a'\boldsymbol{Y}-l'\hat{\beta}+l'\hat{\beta})$$

$$=D(a'\boldsymbol{Y}-l'\hat{\beta})+D(l'\hat{\beta})+2\operatorname{cov}(a'\boldsymbol{Y}-l'\hat{\beta}, l'\hat{\beta}),$$

因为

$$\operatorname{cov}(a'\boldsymbol{Y}-l'\hat{\beta}, l'\hat{\beta})$$

$$=\operatorname{cov}[a'\boldsymbol{Y}-a'X(X'X)^{-1}X'\boldsymbol{Y}, a'X(X'X)^{-1}X'\boldsymbol{Y}]$$

$$=\operatorname{cov}\{a'[I-X(X'X)^{-1}X']\boldsymbol{Y}, a'X(X'X)^{-1}X'\boldsymbol{Y}\}$$

$$=a'[I-X(X'X)^{-1}X']D(\boldsymbol{Y})X(X'X)^{-1}X'a$$

$$=\sigma^2a'[I-X(X'X)^{-1}X']X(X'X)^{-1}X'a=0,$$

所以 $D(a'\boldsymbol{Y})=D(a'\boldsymbol{Y}-l'\hat{\beta})+D(l'\hat{\beta})\geqslant D(l'\hat{\beta})$. ▌

特别 $\hat{\beta}_i$ 是 β_i 的最小方差线性无偏估计, $i=0, 1, 2, \cdots, p$.

性质 3 $D(\hat{\beta})=\sigma^2(X'X)^{-1}$, $D(\hat{\beta}_{(1)})=\sigma^2(\widetilde{X}'\widetilde{X})^{-1}$.

【证】 $D(\hat{\beta})=D[(X'X)^{-1}X'\boldsymbol{Y}]$

$$=(X'X)^{-1}X'D(\boldsymbol{Y})X(X'X)^{-1}$$

$$=\sigma^2(X'X)^{-1}X'X(X'X)^{-1}=\sigma^2(X'X)^{-1}.$$

$$D(\hat{\beta}_{(1)})=D[(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{Y}],$$

由于

$$\widetilde{X}'\widetilde{Y}=(\sum_t(x_{t1}-\bar{x}_1)(Y_t-\bar{Y}),\cdots,\sum_t(x_{tp}-\bar{x}_p)(Y_t-\bar{Y}))$$

$$=(\sum_t(x_{t1}-\bar{x}_1)Y_t,\ \cdots,\ \sum_t(x_{tp}-\bar{x}_p)Y_t)=\widetilde{X}'Y,$$

故 $D(\hat{\beta}_{(1)})=D[(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Y]=\sigma^2(\widetilde{X}'\widetilde{X})^{-1}$. ∎

特别 $p=1$ 时,

$$X'=\begin{pmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_n\end{pmatrix},$$

$$X'X=\begin{pmatrix}n & \sum_t x_t\\ \sum_t x_t & \sum_t x_t^2\end{pmatrix},$$

$$(X'X)^{-1}=\frac{1}{n\sum_t(x_t-\bar{x})^2}\begin{pmatrix}\sum_t x_t^2 & -\sum_t x_t\\ -\sum_t x_t & n\end{pmatrix},$$

$$D(\hat{\beta}_0)=\frac{\sum_t x_t^2}{n\sum_t(x_t-\bar{x})^2}\sigma^2=\frac{\sum_t(x_t-\bar{x})^2+n\bar{x}^2}{n\sum_t(x_t-\bar{x})^2}\sigma^2$$

$$=\left[\frac{1}{n}+\frac{\bar{x}^2}{\sum_t(x_t-\bar{x})^2}\right]\sigma^2,$$

$$D(\hat{\beta}_1)=\frac{n}{n\sum_t(x_t-\bar{x})^2}\sigma^2=\frac{\sigma^2}{\sum_t(x_t-\bar{x})^2},\tag{26}$$

$$\operatorname{cov}(\hat{\beta}_0,\ \hat{\beta}_1)=\frac{-\sum_t x_t}{n\sum_t(x_t-\bar{x})^2}\sigma^2=-\frac{\bar{x}\sigma^2}{\sum_t(x_t-\bar{x})^2}.$$

性质 4 $\hat{\sigma}^2$ 是 σ^2 的无偏估计.

【证】

$$
\begin{aligned}
S(\hat{\beta}) &= (\boldsymbol{Y}-X\hat{\beta})'(\boldsymbol{Y}-X\hat{\beta}) \\
&= (\boldsymbol{Y}-X(X'X)^{-1}X'\boldsymbol{Y})'(\boldsymbol{Y}-X(X'X)^{-1}X'\boldsymbol{Y}) \\
&= \boldsymbol{Y}'[I_n-X(X'X)^{-1}X'][I_n-X(X'X)^{-1}X']\boldsymbol{Y} \\
&= \boldsymbol{Y}'[I_n-X(X'X)^{-1}X']\boldsymbol{Y}.
\end{aligned}
$$

上面最后一个等式的导出，利用了 $I_n-X(X'X)^{-1}X'$ 是幂等阵．此外，注意到 $[I_n-X(X'X)^{-1}X']X=0$，又得

$$
\begin{aligned}
S(\hat{\beta}) &= (\boldsymbol{Y}-X\beta)'[I_n-X(X'X)^{-1}X'](\boldsymbol{Y}-X\beta) \\
&= \boldsymbol{\varepsilon}'[I_n-X(X'X)^{-1}X']\boldsymbol{\varepsilon} \\
&= \operatorname{tr}[I_n-X(X'X)^{-1}X']\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'. \qquad (27)
\end{aligned}
$$

所以，

$$
\begin{aligned}
ES(\hat{\beta}) &= \operatorname{tr}([I_n-X(X'X)^{-1}X']\cdot E\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\
&= \operatorname{tr}([I_n-X(X'X)^{-1}X']\cdot\sigma^2 I_n) \\
&= \sigma^2(\operatorname{tr}I_n-\operatorname{tr}(X'X)^{-1}X'X) \\
&= \sigma^2(n-p-1),
\end{aligned}
$$

$$
E\hat{\sigma}^2 = E\,\frac{S(\hat{\beta})}{n-p-1} = \sigma^2. \quad \blacksquare
$$

现在，假定 $\boldsymbol{Y}$ 是正态随机变量：$\boldsymbol{Y}\sim N_n(X\beta,\ \sigma^2 I_n)$，即 $\boldsymbol{\varepsilon}\sim N_n(0,\ \sigma^2 I_n)$．这时可以进一步导出 $\hat{\beta}$，$\hat{\sigma}^2$ 的概率分布．

性质5 $\hat{\beta}\sim N_{p+1}(\beta,\ \sigma^2(X'X)^{-1})$，

$$
\frac{S(\hat{\beta})}{\sigma^2}\sim\chi^2(n-p-1),
$$

且 $\hat{\beta}$ 与 $\hat{\sigma}^2$ 独立．

【证】 $\hat{\beta}$ 是正态随机变量的线性函数，所以 $\hat{\beta}$ 仍是正态随机变量，再利用性质1和性质3即可得出第一个结论：$\hat{\beta}\sim N_{p+1}(\beta,\ \sigma^2(X'X)^{-1})$．

由于 $X(X'X)^{-1}X'$ 是对称，幂等阵，秩为 $p+1$，故知存在正交阵 C，使

$$X(X'X)^{-1}X'=C'\begin{pmatrix} I_{p+1} & 0 \\ 0 & 0 \end{pmatrix}C,$$

$$I-X(X'X)^{-1}X'=C'\left[I-\begin{pmatrix} I_{p+1} & 0 \\ 0 & 0 \end{pmatrix}\right]C$$

$$=C'\begin{pmatrix} 0 & 0 \\ 0 & I_{n-p-1} \end{pmatrix}C. \qquad (28)$$

令 $\boldsymbol{Z}=C\boldsymbol{\varepsilon}$, 根据第二章第六节例 7,

$$\boldsymbol{Z}\sim N_n(0,\ \sigma^2 I_n) \quad 或 \quad \frac{\boldsymbol{Z}}{\sigma}\sim N_n(0,\ I_n).$$

记 $\boldsymbol{Z}=(Z_1,\ \cdots,\ Z_n)'$, 可知 $Z_1,\ \cdots,\ Z_n$ 是独立的随机变量, 特别, $\frac{Z_1}{\sigma},\ \cdots,\ \frac{Z_n}{\sigma}$ 是独立的标准正态随机变量.

由(27), (28)

$$S(\hat{\beta})=\boldsymbol{\varepsilon}'[I-X(X'X)^{-1}X']\boldsymbol{\varepsilon}$$

$$=\boldsymbol{\varepsilon}'C'\begin{pmatrix} 0 & 0 \\ 0 & I_{n-p-1} \end{pmatrix}C\boldsymbol{\varepsilon}=\boldsymbol{Z}'\begin{pmatrix} 0 & 0 \\ 0 & I_{n-p-1} \end{pmatrix}\boldsymbol{Z}$$

$$=Z_{p+2}^2+\cdots+Z_n^2, \qquad (29)$$

故知 $\frac{S(\hat{\beta})}{\sigma^2}=\left(\frac{Z_{p+2}}{\sigma}\right)^2+\cdots+\left(\frac{Z_n}{\sigma}\right)^2\sim\chi^2(n-p-1)$, 第二结论获证.

最后, 由(28)知

$$\hat{\beta}=(X'X)^{-1}X'\boldsymbol{Y}=(X'X)^{-1}X'[(\boldsymbol{Y}-X\beta)+X\beta]$$

$$=(X'X)^{-1}X\boldsymbol{\varepsilon}+\beta$$

$$=(X'X)^{-1}X'\cdot X(X'X)^{-1}X'\boldsymbol{\varepsilon}+\beta$$

$$=(X'X)^{-1}X'C'\begin{pmatrix} I_{p+1} & 0 \\ 0 & 0 \end{pmatrix}C\boldsymbol{\varepsilon}+\beta$$

$$=(X'X)^{-1}X'C'\begin{pmatrix} I_{p+1} & 0 \\ 0 & 0 \end{pmatrix}\boldsymbol{Z}+\beta$$

$$=(X'X)^{-1}X'C'\begin{pmatrix}Z_1\\ \vdots\\ Z_{p+1}\\ 0\\ \vdots\\ 0\end{pmatrix}+\beta$$

是 $Z_1, \cdots, Z_{p+1}$ 的线性函数. 而由(29)知

$$\hat{\sigma}^2=\frac{S(\hat{\beta})}{n-p-1}=\frac{1}{n-p-1}(Z_{p+2}^2+\cdots+Z_n^2)$$

是 $Z_{p+2}, \cdots, Z_n$ 的函数. 由于 $Z_1, \cdots, Z_n$ 独立,故知 $\hat{\beta}$ 与 $\hat{\sigma}^2$ 独立.

5.3 回归分析中的假设检验问题

这一小节要讨论的问题是用变量 $x_1, x_2, \cdots, x_p$ 及 $\boldsymbol{Y}$ 的 n 次观测值 $(x_{t1}, x_{t2}, \cdots, x_{tp}; y_t)$, $t=1, 2, \cdots, n$, 对(6)中某一个或某几个 β_i 是否为零做出判断. 讨论这个问题的目的: 一是考察 $\boldsymbol{Y}$ 与变量 $x_1, x_2, \cdots, x_p$ 是否有线性联系; 二是建立变量 $\boldsymbol{Y}$ 与变量 $x_1, x_2, \cdots, x_p$ 之间的回归方程时,为了使用方便,希望回归方程中包含的变量少一些. 为此希望把那些对 $\boldsymbol{Y}$ 相对说来不起什么作用的变量从回归方程中剔除掉. 看某一个(或某几个)变量,例如 x_p 对 $\boldsymbol{Y}$ 相对说来是否不起什么作用,就看 $\boldsymbol{Y}$ 对 x_p 的回归系数 β_p 是否为零. 我们将把这个问题作为检验统计假设 H: $\beta_p=0$ 的问题来处理.

1. 检验 H: $\beta_i=0$(i 是 0, 1, 2, $\cdots$, p 中某一数).

从 5.2 性质 1 知, $\hat{\beta}_i$ 是 β_i 的无偏估计,所以 H 真时 $E\hat{\beta}_i=0$, $\hat{\beta}_i^2$ 太大时应否定 H. 记 $(X'X)^{-1}=(c_{ij})$, 取形如

$$\left\{\frac{(n-p-1)\hat{\beta}_i^2}{c_{ii}S(\hat{\beta})}>c\right\}$$

的判别区域，则由 5.2 性质 5 知

$$\frac{\hat{\beta}_i-\beta_i}{\sqrt{c_{ii}}\sigma}\sim N(0,\ 1).$$

在 H 真时，$\hat{\beta}_i^2/c_{ii}\sigma^2$ 是自由度为 1 的 χ^2 变量，又 $S(\hat{\beta})/\sigma^2$ 是自由度为 $n-p-1$ 的 χ^2 变量并且与 $\hat{\beta}_i$ 独立，所以在 H 真时

$$\frac{(n-p-1)\hat{\beta}_i^2}{c_{ii}S(\hat{\beta})}=\frac{\hat{\beta}_i^2}{c_{ii}\sigma^2}\Big/\frac{S(\hat{\beta})}{\sigma^2(n-p-1)}$$

是第一自由度为 1，第二自由度为 $n-p-1$ 的 F 变量．给定信度 α 后，取 $c=F^{\alpha}(1,\ n-p-1)$ 即可．最后得到的检验方法是：若 $(n-p-1)\hat{\beta}_i^2/c_{ii}S(\hat{\beta})>F^{\alpha}(1,\ n-p-1)$ 就拒绝 H，否则接受 H．

2．检验 H：$\beta_1=\beta_2=\cdots=\beta_p=0$．

在 H 真时，(7)变为

$$\boldsymbol{Y}=1_n\beta_0+\boldsymbol{\varepsilon},$$

其中 $1_n=(1,\ 1,\cdots,\ 1)'$ 是 n 维列向量．

$$\hat{\beta}_0=(1_n'1_n)^{-1}1_n'\boldsymbol{Y}=\frac{1}{n}\sum_t y_t=\overline{Y}.$$

剩余平方和 $\qquad S_H=\sum_t(Y_t-\overline{Y})^2.$

在 H 真时(13) $S(\hat{\beta})$ 一般应接近 S_H，若 $S_H-S(\hat{\beta})$ 太大就否定 H．取判别区域为

$$\left\{\frac{n-p-1}{p}\,\frac{S_H-S(\hat{\beta})}{S(\hat{\beta})}>c\right\},$$

因为 $$\frac{S_H}{\sigma^2}=\frac{S(\hat{\beta})}{\sigma^2}+\frac{S_H-S(\hat{\beta})}{\sigma^2},$$

$(S_H-S(\hat{\beta}))/\sigma^2$ 非负，$S(\hat{\beta})/\sigma^2$ 是自由度为 $n-p-1$ 的 χ^2 变量，在 H 真时 S_H/σ^2 是自由度为 $n-1$ 的 χ^2 变量，所以在 H 真时，$(S_H-S(\hat{\beta}))/\sigma^2$ 是自由度为 $(n-1)-(n-p-1)=p$ 的 χ^2 变量，并且与 $S(\hat{\beta})/\sigma^2$ 独立，

$$\frac{n-p-1}{p}\cdot\frac{S_H-S(\hat{\beta})}{S(\hat{\beta})}=\frac{S_H-S(\hat{\beta})}{\sigma^2 p}\bigg/\frac{S(\hat{\beta})}{\sigma^2(n-p-1)}$$

是第一自由度为 p 第二自由度为 $n-p-1$ 的 F 变量．给定信度 α 后取 $c=F^{\alpha}(p,\ n-p-1)$ 即可．又由(23)知

$$S_H-S(\hat{\beta})=\sum_{i=1}^{p}l_{iy}\hat{\beta}_i.$$

最后得到的检验方法是：若

$$\frac{n-p-1}{p}\ \frac{\sum\limits_{i=1}^{p}l_{iy}\hat{\beta}_i}{S(\hat{\beta})}>F^{\alpha}(p,\ n-p-1)$$

就拒绝 H，否则接受 H．

【例 3】 某种水泥在凝固时放出的热量 y(卡/克)可能与下列四种化学成分有关：

x_1: $3CaO\cdot Al_2O_3$ 的成分(%)，

x_2: $3CaO\cdot SiO_2$ 的成分(%)，

x_3: $4CaO\cdot Al_2O_3\cdot Fe_2O_3$ 的成分(%)，

x_4: $2CaO\cdot SiO_3$ 的成分(%)．

实际测得 13 组数据如下：

表 5-16

	x_1	x_2	x_3	x_4	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

在这个例中回归模型是

$$Y_t=\beta_0+\beta_1x_{t1}+\beta_2x_{t2}+\beta_3x_{t3}+\beta_4x_{t4}+\varepsilon_t,$$
$$t=1,\ 2,\ \cdots,\ 13, \tag{30}$$

$$X=\begin{bmatrix}1 & 7 & 26 & 6 & 60\\ 1 & 1 & 29 & 15 & 52\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & 11 & 66 & 9 & 12\\ 1 & 10 & 68 & 8 & 12\end{bmatrix},$$

$$\tilde{X}'\tilde{X}=\begin{bmatrix}415.23 & 251.08 & -372.62 & -290.00\\ 251.08 & 2905.69 & -166.54 & -3041.00\\ -372.62 & -166.54 & 492.31 & 38.00\\ -290.00 & -3041.00 & 38.00 & 3362.00\end{bmatrix},$$

$$\tilde{X}'\tilde{Y}=\begin{pmatrix}775.96\\ 2292.95\\ -618.23\\ -2481.70\end{pmatrix},$$

$$(\tilde{X}'\tilde{X})^{-1}=\begin{bmatrix}0.092763 & 0.085736 & 0.092691 & 0.084504\\ 0.085736 & 0.087607 & 0.087917 & 0.085644\\ 0.092691 & 0.087917 & 0.095255 & 0.086441\\ 0.084504 & 0.085644 & 0.086441 & 0.084076\end{bmatrix}.$$

用5.1的方法总可以求出 $\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4$ 的最小二乘估计,因而得到回归方程

$$y=62.4052+1.5511x_1+0.5101x_2+0.1019x_3$$
$$-0.1441x_4. \tag{31}$$

问题是这个方程是不是真的反映了 Y 和 $x_1,\ x_2,\ x_3,\ x_4$ 之间的线性联系. 如果 Y 和 $x_1,\ x_2,\ x_3,\ x_4$ 之间没有线性联系,即 $\beta_1=\beta_2=\beta_3=\beta_4=0$, 那么使用回归方程(31)也就没有什么意义了. 因此,在求出回归方程(31)后需要检验统计假设

$$H_1:\ \beta_1=\beta_2=\beta_3=\beta_4=0,$$

取信度 $\alpha=0.05$. 本例 $p=4,\ n=13$, 查 F 分布表得

$$F^{0.05}(4,\ 8)=3.84,$$

$$S_{H_1}=\sum_{t=1}^{13}(y_t-\bar{y})^2=2715.76,$$

$$S(\hat{\beta})=\sum_{t=1}^{13}(y_t-\bar{y})^2-\sum_{i=1}^{5}l_{iy}\hat{\beta}_i$$

$$=2715.76-(775.96\times1.5511+2292.95\times0.5101$$

$$-618.23\times0.1019+2481.70\times0.1441)$$

$$=47.86.$$

因为

$$\frac{n-p-1}{p}\ \frac{S_{H_1}-S(\hat{\beta})}{S(\hat{\beta})}=\frac{8}{4}\cdot\frac{2715.76-47.86}{47.86}$$

$$=111.5>3.84,$$

应当拒绝 H_1，不能认为 β_1，β_2，β_3，β_4 都是零，所以回归方程(31)有意义．下面考虑有些变量是否可以从回归方程中去掉．回归方程(31)中 $\hat{\beta}_3$ 的绝对值比较小，这是不是表示相对说来 x_3 和 Y 的线性联系不明显因而可以去掉呢？为此我们检验统计假设

$$H_2:\ \beta_3=0,$$

取信度 $\alpha=0.05$，查 F 分布表得 $F^{0.05}(1,8)=5.32$．由于

$$\frac{(n-p-1)\hat{\beta}_3^2}{c_{33}S(\hat{\beta})}=\frac{8\times(0.1019)^2}{0.095255\times47.86}=0.018<5.32,$$

应当接受 H_2，认为 $\beta_3=0$．现在应当考虑回归模型

$$Y_t=\beta_0+\beta_1x_{t1}+\beta_2x_{t2}+\beta_4x_{t4}+\varepsilon_t,\ t=1,2,\cdots,13.\ (32)$$

对(32)中 β_1，β_2，β_4 是否为零也需要检验，用类似的方法可以得到 $\beta_4=0$ 的判断．最后得到的回归方程是

$$y=52.5773+1.4683x_1+0.6623x_2.$$

习 题 五

1．$Y_t=\alpha+\beta x_t+\varepsilon_t$，$\varepsilon_t\sim N(0,\sigma^2)$，$t=1,2,\cdots,8$ 独立，试根据(x_t,Y_t)的 8 组值

x_t	2.2 −4.1 −0.8 3 1.5 −2.8 4 −2.3
y_t	−4.1 9.2 2.7 −4.6 −2.1 7.1 −7.4 5.3

求α, β的最小二乘估计，σ^2的无偏估计。

2. 对回归模型(32)检验H: $\beta_4=0$。

3. $$Y_t=\beta_0+\beta_1 x_{t1}+\beta_2 x_{t2}+\varepsilon_t,\quad t=1, 2, \cdots, 13.$$
试用表5-16的数据检验
(1) H_1: $\beta_1=0$; (2) H_2: $\beta_2=0$.

4. 有一架天平，称重时有随机误差ε，$E\varepsilon=0$，$D\varepsilon=\sigma^2$。现对实重分别为$\beta_1, \beta_2, \beta_3, \beta_4$的4个东西$A_1, A_2, A_3, A_4$($\beta_1, \beta_2, \beta_3, \beta_4$未知)，按下述办法称量4次：
第1次：A_1, A_2, A_3, A_4都放在天平左盘中，右盘砝码读数为Y_1；
第k次：A_1, A_k放在天平左盘中，其余两个放在右盘中，右盘砝码读数为Y_k，$k=2, 3, 4$(如砝码在左盘中，右盘砝码读数为负)。得
$$Y_1=\beta_1+\beta_2+\beta_3+\beta_4+\varepsilon_1,\quad Y_2=\beta_1+\beta_2-\beta_3-\beta_4+\varepsilon_2,$$
$$Y_3=\beta_1-\beta_2+\beta_3-\beta_4+\varepsilon_3,\quad Y_4=\beta_1-\beta_2-\beta_3+\beta_4+\varepsilon_4.$$
试求$\beta_1, \beta_2, \beta_3, \beta_4$的最小二乘估计及其方差。如果对$A_1, A_2, A_3, A_4$分别单独进行称量需要称量多少次才能得到同样精度(方差)的无偏估计?

第五章小结

根据实际观测到的随机试验的结果，对有关事件的概率或随机变量的分布律、数字特征做出估计或推测是数理统计的任务之一。

本章着重介绍了一些常用的统计推断方法：对未知参数的点估计及区间估计，对未知参数或分布的假设检验，因子试验中的方差分析方法及线性回归分析方法。在介绍方法的过

程中讲了鉴定估计量好坏的准则，假设检验中的两类错误及有关统计量分布的重要定理，掌握这些内容可以对上述统计推断方法有一定程度的理解.

这一章是本书的重要组成部分. 从统计推断方法使用的概念和理论依据看，这一章又是概率论的一个系统应用. 学习这一章后要能正确使用这些方法，正确理解这些方法，同时要回过头来体会前几章中讲的事件，概率，独立等概念的重要性，体会研究随机变量的分布律及其数字特征的重要性.

习题答案

第一章

习题一 **3**. $A\subseteq B$-不正确, B、C互不相容-不正确, $A=B\cup C$-不正确, $B\cap C\subseteq A$-正确. **4**. $A_i\subseteq A$, $i=0, 1, 2, 3, 4, 5$, A、B对立因而也互不相容, A_0, A_1, A_2, A_3, A_4, A_5, B两两互不相容. **5**. (1) "至少有一次没有击中", "前两次没有击中", "恰好接连击中两次"; (2) $A_1\overline{A}_2\overline{A}_3+\overline{A}_1A_2\overline{A}_3+\overline{A}_1\overline{A}_2A_3$, $\overline{A}_1(A_2+A_3)$. **6**. $B_1=A_1+A_2A_3+A_4$, $B_2=[A_1(A_2A_3+A_5)+A_6(A_3+A_2A_5)]A_4$. **7**. (1) $A\overline{B}\overline{C}$; (2) ABC; (3) ABC; (4) $A+B+C$; (5) $AB+BC+CA$; (6) $A\overline{B}\overline{C}+\overline{A}B\overline{C}+\overline{A}\overline{B}C$; (7) $AB\overline{C}+A\overline{B}C+\overline{A}BC$; (8) $\overline{AB+BC+CA}$; (9) $\overline{ABC}$; (10) $\overline{A}\overline{B}\overline{C}$.

习题二 **1**. $\frac{1}{12}$. **2**. $\frac{2}{5}$. **3**. $A_{10}^7/10^7$. **4**. $A_{365}^n/365^n$ ($n\leqslant 365$时); 0($n>365$时). **5**. $C_m^aC_n^b/C_{m+n}^{a+b}$. **6**. $\frac{m}{m+n}$. **7**. (1) 排成一列, $2(n-3)/n(n-1)$($n\geqslant 4$时), 0($n<4$时); (2) 排成一个圆圈, 0($n<4$时), $1/5$($n=6$时), $2/(n-1)$($n\geqslant 4$, $n\neq 6$时). **8**. 如理解题意为任取4个, 任意排好, 则排成4位偶数的概率为41/90; 如理解题意为任取4个,则能排成4位偶数的概率为$1-\frac{C_5^4}{C_{10}^4}$. **9**. A_N^n/N^n ($n\leqslant N$), 0($n>N$). **10**. $(N-1)^n/N^n$. **11**. (1) $C_{n-1}^{N-1}/C_{n+N-1}^{N-1}$; (2) $C_N^mC_{n-1}^{N-m-1}/C_{n+N-1}^{N-1}$; (3) $C_{k+m-1}^{m-1}\cdot C_{(n-k)+(N-m)-1}^{(N-m)-1}/C_{n+N-1}^{N-1}$.

习题三 **1**. 0.121. **2**. $\left(1-\frac{t_0}{T_2-T_1}\right)^2$. **3**. 1/2. **4**. 方格边长不大于$1\frac{1}{9}$. **5**. (1) 0.875; (2) 0.5966; (3) 0.25; (4) 0.75.

习题四 **1**. $a+b$, 0, $1-b$, a, $1-a-b$. **2**. $P(A\overline{B})=r-q$, $P(\overline{A}\overline{B})=1-r$. **10**. $1-\frac{1}{N^n}$. **11**. $1-\frac{1}{2!}+\frac{1}{3!}-\frac{1}{4!}+\cdots+(-1)^{n-1}\frac{1}{n!}$.

习题五 **1**. 0.926. **2**. 0.645. **3**. 0.52. **4**. 1/126. **5**. 25/69, 28/69, 16/69. **6**. 0.998, 0.664.

习题六 **3**. 设三个事件为 A, B, C, 由$(A+B)C\subseteq C$ 可得 $x\geqslant 2x^2$, 在 $[0, 1]$上满足这个条件的最大 x 是 $\frac{1}{2}$. 这个值是可以达到的. 在一个质量分布均匀的正四面体的第一个面上涂红、黄两色, 第二个面上涂红、绿两色, 第三个面上涂黄、绿两色, 第四个面上涂蓝色, 将四面体任意投扔一次, 用 A 表示"有红色面着地", B 表示"有黄色面着地", C 表示"有绿色面着地", 则 A,B,C 的概率都是 $\frac{1}{2}$, 两两独立, 并且 $ABC=\emptyset$. **4**. $P(B)=1/3$.

习题七 **1**. (1) 0.107; (2) 0.268; (3) 0.625. **2**. (1) 0.612; (2) 0.997. **3**. 1/3. **4**. 0.84, 6. **5**. $C_{2n-r}^{n}/2^{2n-r}$.

补充题 **1**. 1/15. **2**. $60/7^6$. **3**. 3/8.

4. (1) $p_1=\frac{(l_1+m_1+n_1)!}{l_1!\,m_1!\,n_1!}\left(\frac{l}{l+m+n}\right)^{l_1}\left(\frac{m}{l+m+n}\right)^{m_1}\left(\frac{n}{l+m+n}\right)^{n_1}$;

(2) $p_2=3!\left(\frac{l}{l+m+n}\right)^{l_1}\left(\frac{m}{l+m+n}\right)^{m_1}\left(\frac{n}{l+m+n}\right)^{n_1}$; (3) $p_3=p_2/3!$.

5. 16/33. **6**. (1) 0.504; (2) 0.432; (3) 0.037. **7**. C_{n+1}^{m}/C_{n+m}^{m}.

8. $C_n^{r-k}C_m^{k-1}(r-1)!(m+1-k)/A_{m+n}^{r}$. **9**. (1) 2/25, (2) 19/75.

10. C_N^n/N^n. **11**. $\sum_{k=1}^{n}C_N^kC_{n-1}^{k-1}/N^n$. **12**. (1) $[k^n-(k-1)^n]/N^n$; (2) $[(N-k+1)^n-(N-k)^n]/N^n$. **13**. $\sum_{k=0}^{N}(-1)^kC_N^k(N-k)^n/N^n$.

14. $\sum_{k=0}^{n-1}(-1)^kC_{n-1}^{k}(n-1-k)^{N-1}/n^{N-1}$. **15**. $P(\text{甲胜})=\alpha/(\alpha+\beta-\alpha\beta)$, $P(\text{乙胜})=(\beta-\alpha\beta)/(\alpha+\beta-\alpha\beta)$. **16**. $P(\text{甲胜})=\alpha^2/(1-2\alpha\beta)$, $P(\text{乙胜})=\beta^2/(1-2\alpha\beta)$. **17**. $P(\text{甲胜})=5/14$, $P(\text{乙胜})=5/14$, $P(\text{丙胜})=2/7$.

18. $(a\leqslant\xi<b)=A_b-A_a$,

$(\xi\leqslant b)=\bigcap_{n=1}^{\infty}A_{b+\frac{1}{n}}$,

$(a<\xi<b)=A_b-\bigcap_{n=1}^{\infty}A_{a+\frac{1}{n}}$,

$(\xi<+\infty)=\bigcup_{n=1}^{\infty}A_n$, $(\xi\geqslant a)=\bar{A}_a$.

第　二　章

习题一　1

$$F(x)=\begin{cases}0, & 当\ x\leqslant 2;\\ \frac{1}{36}, & 当\ 2<x\leqslant 3;\\ \frac{3}{36}, & 当\ 3<x\leqslant 4;\\ \frac{6}{36}, & 当\ 4<x\leqslant 5;\\ \frac{10}{36}, & 当\ 5<x\leqslant 6;\\ \frac{15}{36}, & 当\ 6<x\leqslant 7;\\ \frac{21}{36}, & 当\ 7<x\leqslant 8;\\ \frac{26}{36}, & 当\ 8<x\leqslant 9;\\ \frac{30}{36}, & 当\ 9<x\leqslant 10;\\ \frac{33}{36}, & 当\ 10<x\leqslant 11;\\ \frac{35}{36}, & 当\ 11<x\leqslant 12;\\ 1, & 当\ x>12.\end{cases}$$

2.

$$F(x)=\begin{cases}0, & 当\ x\leqslant 2;\\ \frac{1}{10}, & 当\ 2<x\leqslant 3;\\ \frac{3}{10}, & 当\ 3<x\leqslant 4;\\ \frac{6}{10}, & 当\ 4<x\leqslant 5;\\ 1, & 当\ x>5.\end{cases}$$

3.

$$F_\xi(x)=\begin{cases}0, & 当\ x\leqslant 0;\\ \frac{8}{27}, & 当\ 0<x\leqslant 1;\\ \frac{20}{27}, & 当\ 1<x\leqslant 2;\\ \frac{26}{27}, & 当\ 2<x\leqslant 3;\\ 1, & 当\ x>3.\end{cases}$$

$$F_\eta(x)=\begin{cases}0, & 当\ x\leqslant 1;\\ \frac{1}{3}, & 当\ 1<x\leqslant 2;\\ \frac{2}{3}, & 当\ 2<x\leqslant 3;\\ 1, & 当\ x>3.\end{cases}$$

4. $(\Omega, \mathscr{F}, P)$ 是一个概率空间，若 $\xi(\omega)=c$，则由

$$\{\xi(\omega)<x\}=\begin{cases}\varnothing, & 当\ x\leqslant c;\\ \Omega, & 当\ x>c\end{cases}$$

知$\{\xi(\omega)<x\}\in\mathscr{F}$，$\xi(\omega)$是随机变量.

$$F_\xi(x)=\begin{cases}0, & 当\ x<c;\\ 1, & 当\ x>c.\end{cases}$$

5. $P(|\xi|<1)=F(1)-F(-1+0)$, $P(|\xi-2|<3)=F(5)-F(-1+0)$, $P(2\xi+1>5)=1-F(2+0)$, $P(\xi^2\leqslant4)=F(2+0)-F(-2)$, $P(\xi^3<8)=F(2)$.

习题二 **1**. (1) 1; (2) $1/(e^\lambda-1)$; (3) 1.

2. $F(x)=\begin{cases}0, & 当\ x\leqslant1;\\ \sum\limits_{k=1}^{n}pq^{k-1}, & 当\ n<x\leqslant n+1,\ n=1, 2, \cdots.\end{cases}$

3. $P(\xi=2^n)=1/2^n$, $n=1, 2, \cdots$. **4**. 用ξ, η分别表示甲、乙投篮次数. 若甲先投，$P(\xi=k)=0.76\times0.24^{k-1}$, $k=1, 2, \cdots$, $P(\eta=0)=0.4$, $P(\eta=k)=0.456\times0.24^{k-1}$, $k=1, 2, \cdots$. 若乙先投，$P(\xi=0)=0.6$, $P(\xi=k)=0.304\times0.24^{k-1}$, $k=1, 2, \cdots$, $P(\eta=k)=0.76\times0.24^{k-1}$, $k=1, 2, \cdots$. **5**. $1/6e$. **6**. $[\lambda]$($[x]$表示不超过x的最大整数，下同). **7**. $[np+p]$. **8**. $[(M+1)(n+1)/(N+2)]$. **9**. 至少装103只(计算装101只，102只，103只时有100只合格的概率).

10. $\sum\limits_{k=3}^{500}C_{500}^k\left(\frac{1}{500}\right)^k\left(1-\frac{1}{500}\right)^{500-k}\approx1-\sum\limits_{k=0}^{2}\frac{e^{-1}}{k!}=0.0803$.

习题三 **1**. 0.5, 0.5, 0.6. **2**. 0.1056, 0.0401, 0.8543, 0.7338. **3**. $x_1=-0.59$, $x_2=0$, $x_3=0.43$. **4**. 0.87. **5**. 8/27, 1/27. **6**. (1) $c=1$; (2) 0.4; (3) $p(x)=\begin{cases}0, & 当\ x<0\ 或\ x\geqslant1,\\ 2x, & 当\ 0\leqslant x<1.\end{cases}$

7. (1) $A=1/2$, $B=1/\pi$; (2) $p(x)=1/\pi(1+x^2)$; (3) $\frac{1}{4}+\frac{1}{\pi}\operatorname{arc\,tg}2$. **8**. (1) $c=4$; (2) $a=2^{-\frac{1}{4}}\approx0.84$; (3) $b=0.99^{\frac{1}{4}}\approx0.997$. **9**. (1) $A=k^3/2$; (2) 0.0803;

(3) $F(x)=\begin{cases}0, & 当\ x\leqslant0;\\ 1-\frac{1}{2}(k^2x^2+2kx+2)e^{-kx}, & 当\ x>0.\end{cases}$ **10**. 1/3.

习题四 **1**. $F(b, y)-F(a, y)$, $F(a+0, y)-F(a,y)$, $1-F(a, +\infty)-F(+\infty, b)+F(a, b)$,

$F(a_3, b_5)-F(a_2, b_3)-[F(a_5, b_1)-F(a_2, b_1)]-[F(a_1, b_5)-F(a_1, b_3)]-[F(a_4, b_4)-F(a_4, b_2)-F(a_3, b_4)+F(a_3, b_2)]$.

3. $p_{k_1k_2k_3}=P(\xi_1=k_1, \xi_2=k_2, \xi_3=k_3)=\frac{3!}{k_1!k_2!k_3!}\left(\frac{5}{10}\right)^{k_1}\left(\frac{3}{10}\right)^{k_2}\left(\frac{2}{10}\right)^{k_3}$,

$k_1, k_2, k_3\geqslant 0, k_1+k_2+k_3=3$.

$p_{003}=0.008$, $p_{030}=0.027$, $p_{300}=0.125$, $p_{012}=0.036$, $p_{021}=0.054$, $p_{102}=0.06$, $p_{201}=0.15$, $p_{120}=0.135$, $p_{210}=0.225$, $p_{111}=0.18$.

5. (1) ξ有分布列$\begin{pmatrix}0 & 1 & 2 & 3 & 4 & 5\\ 0.03 & 0.08 & 0.16 & 0.21 & 0.24 & 0.28\end{pmatrix}$,

η有分布列$\begin{pmatrix}0 & 1 & 2 & 3\\ 0.25 & 0.26 & 0.25 & 0.24\end{pmatrix}$;

(2) $\begin{pmatrix}0 & 1 & 2 & 3\\ \frac{5}{21} & \frac{5}{21} & \frac{5}{21} & \frac{6}{21}\end{pmatrix}$;

(3) $\begin{pmatrix}0 & 1 & 2 & 3 & 4 & 5\\ \frac{1}{26} & \frac{2}{26} & \frac{4}{26} & \frac{5}{26} & \frac{6}{26} & \frac{8}{26}\end{pmatrix}$.

6. (1) $A=2$, (2) $F(x, y)=\begin{cases}(1-e^{-2x})(1-e^{-y}), & 当x>0, y>0,\\ 0, & 当x\leqslant 0或y\leqslant 0;\end{cases}$

(3) $F_\xi(x)=\begin{cases}1-e^{-2x}, & 当x>0,\\ 0, & 当x\leqslant 0;\end{cases}$ $F_\eta(y)=\begin{cases}1-e^{-y}, & 当y>0,\\ 0, & 当y\leqslant 0;\end{cases}$

(4) $f_{\xi|\eta}(x|y)=\begin{cases}2e^{-2x}, & 当x>0,\\ 0, & 当x\leqslant 0;\end{cases}$ (5) $1-e^{-4}$.

7. (1) 5/6; (2) 1/2; (3) 7/72, (4) 17/24.

习题五 **1**. (1) 独立; (2) 不独立. **2**. (1) 独立; (2) 不独立.

3. $f(x, y)=f_1(x)f_2(y)=\begin{cases}e^{-(x+y)}, & 当x>0, y>0;\\ 0, & 当x\leqslant 0或y\leqslant 0.\end{cases}$

$f_{\xi|\eta}(x|y)=f_1(x)$, $f_{\eta|\xi}(y|x)=f_2(y)$. **4**. 不独立.

习题六 **1**. η的分布列是$\begin{pmatrix}0 & 1 & 4\\ 0.3 & 0.4 & 0.3\end{pmatrix}$,

ζ的分布列是$\begin{pmatrix}-5 & -3 & -1 & 1 & 3\\ 0.15 & 0.2 & 0.3 & 0.2 & 0.15\end{pmatrix}$,

ψ的分布列是$\begin{pmatrix} 1 & 2 & 3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$.

2. ζ的分布列是$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.3 & 0.18 & 0.32 & 0.12 & 0.08 \end{pmatrix}$,

ψ的分布列是$\begin{pmatrix} 0 & 2 & 4 \\ 0.8 & 0.12 & 0.08 \end{pmatrix}$.

7. $p(x)=\begin{cases} \dfrac{2}{\pi(b-a)}\left(\dfrac{6x}{\pi}\right)^{\frac{2}{3}}, & \text{当}\quad \dfrac{\pi a^3}{6}<x\leqslant\dfrac{\pi b^3}{6}; \\ 0, & \text{其它}. \end{cases}$

8. $p_{\eta}(x)=\begin{cases} \dfrac{2}{m^{3/2}\sigma^3\sqrt{\pi}}x^{\frac{1}{2}}e^{-\frac{x}{m\sigma^2}}, & \text{当}\quad x>0; \\ 0, \quad \text{当}\quad x\leqslant 0. \end{cases}$

10. $p(x)=\begin{cases} 0, & \text{当 } x\leqslant 0 \text{ 时}, \\ \dfrac{x^2}{2(1-x)^4}, & \text{当 } x>0 \text{ 时}. \end{cases}$ **12**. 独立.

14. (1) 对任何自然数 k, l,

$$P(\zeta=k,\ \xi=l)=\begin{cases} 0, & \text{当 } k<l \text{ 时}, \\ pq^{l-1}(1-q^l), & \text{当 } k=l \text{ 时}, \\ p^2q^{k+l-2}, & \text{当 } k>l \text{ 时}; \end{cases}$$

(2) $P(\zeta=k)=2pq^{k-1}-2pq^{2(k-1)}+p^2q^{2(k-1)}$, $k=1, 2, \cdots$;

(3) $$P(\xi=l|\zeta=k)=\begin{cases} 0, & \text{当 } l>k \text{ 时}, \\ \dfrac{1-q^k}{2(1-q^{k-1})+pq^{k-1}}, & \text{当 } l=k \text{ 时}, \\ \dfrac{pq^{l-1}}{2(1-q^{k-1})+pq^{k-1}}, & \text{当 } l<k \text{ 时}. \end{cases}$$

补充题

6. $P(z)=\begin{cases} 0, & \text{当 } z\leqslant 0 \quad \text{或} \quad z>l^n; \\ \dfrac{2}{nl^2}(l-z^{\frac{1}{n}})z^{\frac{1}{n}-1}, & \text{当 } 0<z\leqslant l^n. \end{cases}$

8. $F_{\eta_2}(x)=\begin{cases} 0, & \text{当 } x\leqslant a \text{ 时}, \\ \left(\dfrac{x-a}{b-a}\right)^n, & \text{当 } a<x\leqslant b \text{ 时}, \\ 1, & \text{当 } b<x \text{ 时}; \end{cases}$

$$p_{\eta_1}(x)=\begin{cases}0, & \text{当 } x\leqslant a \text{ 或 } x>b \text{ 时},\\ \dfrac{n(x-a)^{n-1}}{(b-a)^n}, & \text{当 } a<x\leqslant b \text{ 时};\end{cases}$$

$$F_{\eta_n}(x)=\begin{cases}0, & \text{当 } x\leqslant a \text{ 时},\\ 1-\left(\dfrac{b-x}{b-a}\right)^n, & \text{当 } a<x\leqslant b \text{ 时},\\ 1, & \text{当 } b<x \text{ 时};\end{cases}$$

$$p_{\eta_n}(x)=\begin{cases}0, & \text{当 } x\leqslant a \text{ 或 } x>b \text{ 时},\\ \dfrac{n(b-x)^{n-1}}{(b-a)^n}, & \text{当 } a<x\leqslant b \text{ 时}.\end{cases}$$

9. (η_1, η_n)的联合分布密度为

$$p(x, y)=\begin{cases}\dfrac{n(n-1)}{(b-a)^n}(y-x)^{n-2}, & \text{当 } a<x<y<b,\\ 0, & \text{其它}.\end{cases}$$

10. $\eta_1+\eta_n$ 的分布密度是

$$p(x)=\begin{cases}0, & \text{当 } x\leqslant 2a, \text{ 或 } x>2b \text{ 时},\\ \dfrac{n(x-2a)^{n-1}}{2(b-a)^n}, & \text{当 } 2a<x\leqslant a+b \text{ 时},\\ \dfrac{n(2b-x)^{n-1}}{2(b-a)^n}, & \text{当 } a+b<x\leqslant 2b \text{ 时}.\end{cases}$$

第 三 章

习题一 **1**. $E\xi=2$, $E|\xi|^3=53$. **2**. $E\xi=\beta$, $E\xi^2=2\alpha^2+\beta^2$. **3**. $E\xi=\dfrac{2a}{\sqrt{\pi}}$, $E\xi^2=\dfrac{3a^2}{2}$. **4**. $E\xi=3\dfrac{1}{2}$. **5**. $E\xi=\dfrac{n}{2}+2$ [提示：设在倒数第 η 次，倒数第二次取得红球，则 $E\xi=E\eta$. ξ 和 η 还有什么关系?] **6**. $2l^n/(n+1)(n+2)$. **8**. k 为奇数时，$E(\xi-a)^k=0$; k 为偶数时，$E(\xi-a)^k=1\cdot 3\cdots(k-1)\sigma^k$.

习题二 **2**. (1) $r_1=0$, $r_2=0$. (2) $r_1=2^{\frac{3}{2}}/n^{\frac{1}{2}}$, $r_2=12/n$.
4. $a+\dfrac{\sigma}{\sqrt{\pi}}$.

习题三 **1**. $(b-a)^2/12$. **2**. $2\alpha^2$. **3**. $\left(\dfrac{3}{2}-\dfrac{4}{\pi}\right)a^2$. **4**. $D\xi=\sigma_1^2$, $D\eta=\sigma_2^2$. **7**. $E\xi=e^{\alpha+\frac{\beta^2}{2}}$, $D\xi=e^{2\alpha+\beta^2}(e^{\beta^2}-1)$.

习题四 1. (η_1, η_2)服从参数为$\left(na,\ ma,\ n\sigma^2,\ m\sigma^2\sqrt{\frac{m}{n}}\right)$的二维正态分布. 4. $\operatorname{cov}(\xi, \eta)=0$.

习题五 1. $\frac{1}{(b-a)i}[(\sin tb-\sin ta)+i(\cos ta-\cos tb)]$.

2. $(pe^{it}+q)^n$, $q=1-p$. 3. $(1-it)^{-1}$. 4. $\left(1-\frac{it}{b}\right)^{-1}$. 5. $E\eta=\frac{1}{b}$, $D\eta=\frac{1}{b^2}$. 7. (1) $E\xi=2$; (2) $D\xi=1$; (3) 5/8; (4) 5/16.

补充题 1. (iii) $P_\xi(t)=p^r(1-qt)^{-r}$, $E\xi=rq/p$, $D\xi=rq/p^2$. 2. 9.

3. $E\mu^3=n(n-1)(n-2)p^3+3n(n-1)p^2+np$, $E\mu^4=n(n-1)(n-2)\times(n-3)p^4+6n(n-1)(n-2)p^3+7n(n-1)p^2+np$.

4. (i) $E\eta_1=\frac{na+b}{n+1}$, $E\eta_2=\frac{a+nb}{n+1}$;

(ii) $D\eta_1=D\eta_n=\frac{n(b-a)^2}{(n+2)(n+1)^2}$; (iii) $2(b-a)^2/(n+1)(n+2)$.

5. $E\xi=(2n+1)/3$, $D\xi=(n+2)(n-1)/9$. 6. 参数为 N, M, n 的超几何分部数学期望是 nM/N, 方差是$\left(\frac{N-n}{N-1}\right)\frac{nM}{N}\left(1-\frac{M}{N}\right)$.

第 五 章

习题一 4. 提示: $x_i-y_i\sim N(a-b,\ \sigma_1^2+\sigma_2^2)$, $i=1, 2, \cdots, n$.

5. 提示: $\sum_{i=1}^{n}(x_i-\bar{x})^2/\sigma^2\sim\chi^2(n-1)$.

习题二 3. $\hat{\theta}=\bar{x}$. 4. $\hat{\theta}=\frac{1}{2}[\max_{1\le i\le n}\{x_i\}+\min_{1\le i\le n}\{x_i\}]$.

5. $\hat{\theta}=n\Big/\left|\sum_{i=1}^{n}\ln x_i\right|$.

6. $E\xi=\max_{1\le i\le n}\{x_i\}/2$, $D\xi=(\max_{1\le i\le n}\{x_i\})^2/12$.

7. $D\left(\frac{n+1}{n}\max_{1\le i\le n}\{x_i\}\right)\le D\left((n+1)\min_{1\le i\le n}\{x_i\}\right)$.

8. $D\left[\frac{1}{2}\left(\max_{1\le i\le n}\{x_i\}+\min_{1\le i\le n}\{x_i\}\right)\right]\le D\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)$.

9. $k_1=\frac{1}{k+1}$, $k_2=\frac{k}{k+1}$.

11. $P\left[(\bar{x}-\bar{y})-t_\alpha\sqrt{\frac{s_x^2+s_y^2-2s_{xy}}{n-1}}<a-b<(\bar{x}-\bar{y})+t_\alpha\sqrt{\frac{s_x^2+s_y^2-2s_{xy}}{n-1}}\right]$

$=1-\alpha$, t_α 由 $\int_{t_\alpha}^{+\infty} p_{t_{n-1}}(x)dx=\frac{\alpha}{2}$ 确定, $p_{t_{n-1}}(x)$ 是自由度为 $n-1$ 的 t 分布的分布密度. (2.35, 3.49)

习题三 **2**. (1) 判别区域为 $\left\{\sum_{i=1}^{n} x_i<c\right\}$, c 由

$$\int_0^c p_{\chi^2(2n)}(x)dx=0.05$$

确定, $p_{\chi^2(2n)}(x)$ 是自由度为 $2n$ 的 χ^2 分布的分布密度;

(2) 判别区域为 $\left\{\sum_{i=1}^{n} x_i>c\right\}$, c 由

$$\int_c^{+\infty} p_{\chi^2(2n)}(x)dx=0.10$$

确定, $p_{\chi^2(2n)}(x)$ 是自由度为 $2n$ 的 χ^2 分布的分布密度.

附表1 泊松分布

本表给出了

$$p_k=\frac{\lambda^k}{k!}e^{-\lambda}$$

的数值.

k \ λ	0.1	0.2	0.3	0.4	0.5	0.6
0	0.904837	0.818731	0.740818	0.670320	0.606531	0.548812
1	0.090484	0.163746	0.222245	0.268128	0.303265	0.329287
2	0.004524	0.016375	0.033337	0.053626	0.075816	0.098786
3	0.000151	0.001091	0.003334	0.007150	0.012636	0.019757
4	0.000004	0.000055	0.000250	0.000715	0.001580	0.002964
5		0.000002	0.000015	0.000057	0.000158	0.000356
6			0.000001	0.000004	0.000013	0.000035
7					0.000001	0.000003

k \ λ	0.7	0.8	0.9	1.0	2.0	3.0
0	0.496585	0.449329	0.406570	0.367879	0.135335	0.049787
1	0.347610	0.359463	0.365913	0.367879	0.270671	0.149361
2	0.121663	0.143785	0.164661	0.183940	0.270671	0.224042
3	0.028388	0.038343	0.049398	0.061313	0.180447	0.224042
4	0.004968	0.007669	0.011115	0.015328	0.090224	0.168031
5	0.000695	0.001227	0.002001	0.003066	0.036089	0.100819
6	0.000081	0.000164	0.000300	0.000511	0.012030	0.050409
7	0.000008	0.000019	0.000039	0.000073	0.003437	0.021604
8		0.000002	0.000004	0.000009	0.000859	0.008101
9				0.000001	0.000191	0.002701
10					0.000038	0.000810
11					0.000007	0.000221
12					0.000001	0.000055
13						0.000013
14						0.000003
15						0.000001

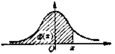
(续表)

k \ λ	4.0	5.0	6.0	7.0	8.0	9.0
0	0.018316	0.006738	0.002479	0.000912	0.000335	0.000123
1	0.073263	0.033690	0.014873	0.006383	0.002684	0.001111
2	0.146525	0.084224	0.044618	0.022341	0.010735	0.004998
3	0.195367	0.140374	0.089235	0.052129	0.028626	0.014994
4	0.195367	0.175467	0.133853	0.091226	0.057252	0.033737
5	0.156293	0.175467	0.160623	0.127717	0.091604	0.060727
6	0.104194	0.146223	0.160623	0.149003	0.122138	0.091090
7	0.059540	0.104445	0.137677	0.149003	0.139587	0.117116
8	0.029770	0.065278	0.103258	0.130377	0.139587	0.131756
9	0.013231	0.036266	0.068838	0.101405	0.124077	0.131756
10	0.005292	0.018133	0.041303	0.070983	0.099262	0.118580
11	0.001925	0.008242	0.022529	0.045171	0.072190	0.097020
12	0.000642	0.003434	0.011262	0.026350	0.048127	0.072765
13	0.000197	0.001321	0.005199	0.014188	0.029616	0.050376
14	0.000055	0.000472	0.002228	0.007094	0.016924	0.032384
15	0.000015	0.000157	0.000891	0.003311	0.009026	0.019431
16	0.000004	0.000049	0.000334	0.001448	0.004513	0.010930
17	0.000001	0.000014	0.000118	0.000596	0.002124	0.005786
18		0.000004	0.000039	0.000232	0.000944	0.002893
19		0.000001	0.000012	0.000085	0.000397	0.001370
20			0.000004	0.000030	0.000159	0.000617
21			0.000001	0.000010	0.000061	0.000264
22				0.000003	0.000022	0.000108
23				0.000001	0.000008	0.000042
24					0.000003	0.000016
25					0.000001	0.000006
26						0.000002
27						0.000001

附表2 标准正态分布

本表给出了

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt \quad (x\geqslant 0)$$



的数值.

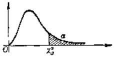
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0.1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0.2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0.3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0.4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0.5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0.6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0.7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0.8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0.9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1.0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1.1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1.2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1.3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1.4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1.5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1.6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1.7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1.8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1.9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2.0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2.1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2.2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2.3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2.4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2.5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2.6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2.7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2.8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2.9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	0.99865	99903	99931	99952	99966	99977	99984	99989	99993	99995

附表3 χ^2 分布

本表给出了满足

$$P(\chi^2>\chi^2_\alpha)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_{\chi^2_\alpha}^{+\infty}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx$$



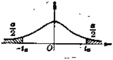
的 χ^2_α 值.(n 是自由度)

n	α									
	0.99	0.98	0.95	0.90	0.50	0.10	0.05	0.02	0.01	0.001
	χ^2_α									
1	0.000	0.001	0.004	0.016	0.455	2.71	3.84	5.41	6.64	10.83
2	0.020	0.040	0.103	0.211	1.386	4.61	5.99	7.82	9.21	13.82
3	0.115	0.185	0.352	0.584	2.366	6.25	7.82	9.84	11.34	16.27
4	0.297	0.429	0.711	1.064	3.357	7.78	9.49	11.67	13.28	18.47
5	0.554	0.752	1.145	1.610	4.351	9.24	11.07	13.39	15.09	20.52
6	0.872	1.134	1.635	2.204	5.35	10.65	12.59	15.03	16.81	22.46
7	1.239	1.564	2.167	2.833	6.35	12.02	14.07	16.62	18.48	24.32
8	1.646	2.032	2.733	3.490	7.34	13.36	15.51	18.17	20.09	26.13
9	2.088	2.532	3.325	4.168	8.34	14.68	16.92	19.68	21.67	27.88
10	2.558	3.059	3.940	4.865	9.34	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.57	5.58	10.34	17.28	19.68	22.62	24.73	31.26
12	3.57	4.18	5.23	6.30	11.34	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	12.34	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	13.34	21.06	23.69	26.87	29.14	36.12
15	5.23	5.99	7.26	8.55	14.34	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	15.34	23.54	26.30	29.63	32.00	39.25
17	6.41	7.26	8.67	10.09	16.34	24.77	27.59	31.00	33.41	40.79
18	7.02	7.91	9.39	10.87	17.34	25.99	28.87	32.35	34.81	42.31
19	7.63	8.57	10.12	11.65	18.34	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	19.34	28.41	31.41	35.02	37.57	45.32
21	8.90	9.91	11.59	13.24	20.34	29.61	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	21.34	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	22.34	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	23.34	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	24.34	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	25.34	35.56	38.89	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	26.34	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	27.34	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	28.34	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	29.34	40.26	43.77	47.96	50.89	59.70

附表 4 t 分 布

本表给出了满足

$$P(|t|>t_\alpha)=1-\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}$$



$$\times\int_{-t_\alpha}^{t_\alpha}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx=\alpha$$

的 t_α 值.(n 是自由度)

n	α = 0.10	0.05	0.02	0.01	0.001
	t_α				
1	6.31	12.71	31.82	63.66	636.62
2	2.92	4.30	6.97	9.93	31.60
3	2.35	3.18	4.54	5.84	12.94
4	2.13	2.78	3.75	4.60	8.61
5	2.02	2.57	3.37	4.03	6.86
6	1.94	2.45	3.14	3.71	5.96
7	1.90	2.37	3.00	3.50	5.41
8	1.86	2.31	2.90	3.36	5.04
9	1.83	2.26	2.82	3.25	4.78
10	1.81	2.23	2.76	3.17	4.59

(续表)

n	α				
	0.10	0.05	0.02	0.01	0.001
	t_α				
11	1.80	2.20	2.72	3.11	4.44
12	1.78	2.18	2.68	3.06	4.32
13	1.77	2.16	2.65	3.01	4.22
14	1.76	2.15	2.62	2.98	4.14
15	1.75	2.13	2.60	2.95	4.07
16	1.75	2.12	2.58	2.92	4.02
17	1.74	2.11	2.57	2.90	3.97
18	1.73	2.10	2.55	2.88	3.92
19	1.73	2.09	2.54	2.86	3.88
20	1.73	2.09	2.53	2.85	3.85
21	1.72	2.08	2.52	2.83	3.82
22	1.72	2.07	2.51	2.82	3.79
23	1.71	2.07	2.50	2.81	3.77
24	1.71	2.06	2.49	2.80	3.75
25	1.71	2.06	2.48	2.79	3.73
26	1.71	2.06	2.48	2.78	3.71
27	1.70	2.05	2.47	2.77	3.69
28	1.70	2.05	2.47	2.76	3.67
29	1.70	2.04	2.46	2.76	3.66
30	1.70	2.04	2.46	2.75	3.65
40	1.68	2.02	2.42	2.70	3.55
60	1.67	2.00	2.39	2.66	3.46
120	1.66	1.98	2.36	2.62	3.37
∞	1.65	1.96	2.33	2.58	3.29

附表5 F 分布

本表给出了满足

$$P(F>F_\alpha)=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\int_{F_\alpha}^{+\infty}x^{\frac{m-2}{2}}(n+mx)^{-\frac{m+n}{2}}dx=\alpha$$



的 F_α 值.（m是第一自由度，n 是第二自由度.）

$\alpha=0.25$

n	m=1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.76	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.46	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.64	1.63	1.62	1.62	1.61	1.59	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.54	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.50	1.48

(续表)

n	$\alpha=0.25$														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.47	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.44	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.42	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.40	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.38	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.36	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.35	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.34	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.33	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.32	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.31	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.30	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.30	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.29	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.28	1.25
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.26	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.22	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.19	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.16	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.12	1.00

(续表)

n	$\alpha=0.05$														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	252.2	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.43	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.62	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.49	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.38	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.30	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.22	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.16	2.07

(续表)

n	α=0.05														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.11	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.06	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.02	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	1.98	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	1.95	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	1.92	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	1.89	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	1.86	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.84	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.82	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.74	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.64	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.53	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.32	1.00

（续表）

n	$\alpha=0.01$														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6313	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.32	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.65	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.20	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.06	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.82	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.03	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.48	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.08	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	3.78	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.54	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.34	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.18	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.05	2.87

(续表)

n	$\alpha=0.01$														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	2.93	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	2.83	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.75	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.67	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.61	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.55	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.50	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.45	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.40	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.36	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.21	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.02	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	1.84	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.66	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.47	1.00

附表6 柯尔莫哥洛夫-斯米尔诺夫分布

本表给出了

$$K(z)=\sum_{l=-\infty}^{+\infty}(-1)^l e^{-2l^2z^2}$$

的数值.

z	$K(z)$	z	$K(z)$	z	$K(z)$	z	$K(z)$	z	$K(z)$
0.32	0.0000	0.58	0.1104	0.84	0.5194	1.10	0.8223	1.36	0.9505
0.33	0.0001	0.59	0.1228	0.85	0.5347	1.11	0.8299	1.37	0.9531
0.34	0.0002	0.60	0.1357	0.86	0.5497	1.12	0.8374	1.38	0.9556
0.35	0.0003	0.61	0.1492	0.87	0.5645	1.13	0.8445	1.39	0.9580
0.36	0.0005	0.62	0.1662	0.88	0.5791	1.14	0.8514	1.40	0.9603
0.37	0.0008	0.63	0.1778	0.89	0.5933	1.15	0.8580	1.41	0.9625
0.38	0.0013	0.64	0.1927	0.90	0.6073	1.16	0.8644	1.42	0.9646
0.39	0.0019	0.65	0.2080	0.91	0.6209	1.17	0.8706	1.43	0.9665
0.40	0.0028	0.66	0.2236	0.92	0.6343	1.18	0.8765	1.44	0.9684
0.41	0.0040	0.67	0.2396	0.93	0.6473	1.19	0.8823	1.45	0.9702
0.42	0.0055	0.68	0.2558	0.94	0.6601	1.20	0.8877	1.46	0.9718
0.43	0.0074	0.69	0.2722	0.95	0.6725	1.21	0.8930	1.47	0.9734
0.44	0.0097	0.70	0.2888	0.96	0.6846	1.22	0.8981	1.48	0.9750
0.45	0.0126	0.71	0.3055	0.97	0.6964	1.23	0.9030	1.49	0.9764
0.46	0.0160	0.72	0.3223	0.98	0.7079	1.24	0.9076	1.50	0.9778
0.47	0.0200	0.73	0.3391	0.99	0.7191	1.25	0.9121	1.51	0.9791
0.48	0.0247	0.74	0.3560	1.00	0.7300	1.26	0.9164	1.52	0.9803
0.49	0.0300	0.75	0.3728	1.01	0.7406	1.27	0.9206	1.53	0.9815
0.50	0.0361	0.76	0.3896	1.02	0.7508	1.28	0.9245	1.54	0.9826
0.51	0.0428	0.77	0.4064	1.03	0.7608	1.29	0.9283	1.55	0.9836
0.52	0.0503	0.78	0.4230	1.04	0.7704	1.30	0.9319	1.56	0.9846
0.53	0.0585	0.79	0.4395	1.05	0.7798	1.31	0.9354	1.57	0.9855
0.54	0.0675	0.80	0.4559	1.06	0.7889	1.32	0.9387	1.58	0.9864
0.55	0.0772	0.81	0.4720	1.07	0.7976	1.33	0.9418	1.59	0.9873
0.56	0.0876	0.82	0.4880	1.08	0.8061	1.34	0.9449	1.60	0.9880
0.57	0.0987	0.83	0.5038	1.09	0.8143	1.35	0.9478	1.61	0.9888

（续表）

z	$K(z)$	z	$K(z)$	z	$K(z)$	z	$K(z)$	z	$K(z)$
1.62	0.9895	1.76	0.9959	1.90	0.9985	2.04	0.9995	2.18	0.9999
1.63	0.9902	1.77	0.9962	1.91	0.9986	2.05	0.9996	2.19	0.9999
1.64	0.9908	1.78	0.9965	1.92	0.9987	2.06	0.9996	2.20	0.9999
1.65	0.9914	1.79	0.9967	1.93	0.9988	2.07	0.9996	2.21	0.9999
1.66	0.9919	1.80	0.9969	1.94	0.9989	2.08	0.9996	2.22	0.9999
1.67	0.9924	1.81	0.9971	1.95	0.9990	2.09	0.9997	2.23	0.9999
1.68	0.9929	1.82	0.9973	1.96	0.9991	2.10	0.9997	2.24	0.9999
1.69	0.9934	1.83	0.9975	1.97	0.9991	2.11	0.9997	2.25	0.9999
1.70	0.9938	1.84	0.9977	1.98	0.9992	2.12	0.9997	2.26	0.9999
1.71	0.9942	1.85	0.9979	1.99	0.9993	2.13	0.9998	2.27	0.9999
1.72	0.9946	1.86	0.9980	2.00	0.9993	2.14	0.9998	2.28	0.9999
1.73	0.9950	1.87	0.9981	2.01	0.9994	2.15	0.9998	2.29	0.9999
1.74	0.9953	1.88	0.9983	2.02	0.9994	2.16	0.9998	2.30	0.9999
1.75	0.9956	1.89	0.9984	2.03	0.9995	2.17	0.9998	2.31	1.0000